UNIVERSITEIT GENT

VIB
FLANDERS INTERUNIVERSITY INSTITUTE FOR BIOTECHNOLOGY

Faculteit Wetenschappen
Vakgroep Moleculaire Genetica

Improvement of the methionine content of seeds through modification of the *Phaseolus vulgaris* seed protein arcelin-5a

Janniek De Clercq

proefschrift voorgelegd tot het bekomen
van de graad van Doctor in de Wetenschappen,
richting Biotechnologie

Academiejaar 2001-2002

promotoren:
Prof. Dr. G. Angenon
Prof. Dr. A. Depicker
Prof. em. Dr. M. Van Montagu

table of contents

GENERAL INTRODUCTION

Protein is considered as an elementary constituent of nutrition for humans and animals, as reflected in the Greek origin of its name, *proteios*, meaning primary. For infants, children, and young animals, proteins are required for growth and development, whereas for adults, a regular supply of proteins is needed to replace routine losses. Dietary proteins come from milk, egg, fish, meat, cereals, legumes and other sources.

The nutritional quality of a protein can be measured by a variety of methods. Basically, it is the relative amount and availability of essential amino acids in the dietary protein that determine its nutritional value. Humans and monogastric animals cannot synthesize the amino acids isoleucine, leucine, lysine, methionine, tryptophan, phenylalanine, threonine, and valine in their body. These amino acids are therefore named essential amino acids and they must be obtained from the diet. In order to be used efficiently from ingested protein, they must be present simultaneously and in balanced amounts – the ones that are present in amounts in excess of the amounts defined by the first limiting amino acid are catabolized for energy production.

Plant seeds represent the major source of proteins in the human diet. They are consumed either directly, or indirectly after transformation to other proteins by fermentation or feeding to animals. In 1990, the world protein supply was 137 million ton, of which 65 % was obtained directly from plants and 35 % through animal protein production (FAO, 1990). Compared to meat, plant protein is much more economical to produce, store and transport. When used as single source of dietary protein for humans and monogastric livestock, however, most plant proteins are nutritionally incomplete due to their deficiency in several of the essential amino acids. Unbalanced amino acid composition can have severe detrimental consequences for people eating a vegetarian diet of limited diversity. As a consequence, the physical and mental development of children up to 4 year might be irreversibly retarded by deficiencies of essential amino acids in their diet. In addition, amino acid imbalance in livestock feed contributes to environmental pollution since it entails an inefficient conversion of plant protein nitrogen to animal protein nitrogen that results in a high level of nitrogen excretion. The above-mentioned facts provide a good reason to improve the nutritional quality of plant proteins by balancing the amino acid content.

The legume genus *Phaseolus*, of which *P. vulgaris* is economically the most important species, is a primary food source providing energy and proteins for millions of people, mostly in Latin America, Africa and India. The world production of *Phaseolus* beans is more than 20 million ton per year (Parker, 1995; Singh, 1999). Their amino acids composition, however, is unbalanced with reference to the needs of humans as well as domesticated monogastric animals because they contain low amounts of the sulfur-

containing amino acids. Compared to the FAO reference protein containing 2.5 % methionine and cysteine (Scrimshaw et al., 1986), *Phaseolus* beans show a relative shortage of these amino acids, with no more than 1,9 % of their content being methionine and cysteine (FAO, 1970).

Humans and non-ruminant animals can convert methionine to cysteine. Therefore, increasing the methionine content of legume seed proteins would overcome this deficiency and help to solve the problems caused by monotonous legume protein diets.

Breeding efforts to enhance the methionine level in legume seeds have so far been unsuccessful. As an alternative approach, gene engineering offers a direct method for manipulating the amino acid composition of seed storage proteins. Several possible strategies to modify the sulfur-containing amino acid content of seeds are described in literature.

The goal of the work presented here is to contribute to the development of a method to improve the methionine content of legume seeds in general and *Phaseolus* beans in particular. The strategy followed in this work is modification and seed-specific expression of the gene coding for the *Phaseolus vulgaris* seed protein arcelin-5a.

LITERATURE SURVEY

The methionine pathway

Methionine is the only sulfur-containing amino acid that is essential for mammals and must therefore be derived entirely from the diet. Plants and micro-organisms, however, can synthesize methionine *de novo* after the initial steps of the aspartate pathway together with inorganic sulfur assimilation and cysteine synthesis.

Inorganic sulfur assimilation and cysteine synthesis

Pathway

Sulfur is taken up by plants in its inorganic sulfate form (SO_4^{2-} ion) from the soil through the root system or to a lesser extent as volatile sulfur compounds from the air. The sulfate uptake from external environments and allocation between plant organs is mediated by specific sulfate transporters localized in the plasma membrane. Several functionally different sulfate transporters were postulated and have now been cloned from a number of plant species (Höfgen et al., 2001 and references therein). Transport throughout the plant is performed mostly in the unmetabolized form; SO_4^{2-} is a major anionic component of vascular and vacuolar sap. Besides the plasma membrane transporters that mediate uptake and long distance transport of sulfate, an intracellular transport system for sulfate is also thought to be present. Once inside plant cells, sulfate is transported into the vacuole for storage or into chloroplasts where reduction and activation of the chemically relatively inert sulfate molecule takes place. The reduction from SO_4^{2-} to S^{2-} is performed in a multi-step pathway in which eight electrons are added:

$$SO_4^{2-} + ATP + 8\ e^- + 8\ H^+ \rightarrow S^{2-} + 4\ H_2O + AMP + PPi$$

For this reduction, the sulfate ion is initially activated to its appropriate forms and then reduced by electrons generated through photosystem I or NADPH in photosynthetic plant cells and non-green tissues, respectively. The pathway of activation and reduction used in plants is depictured in figure 1.1. First, **ATP sulfurylase** (ATP-S) activates sulfate by converting it to adenosine 5'-phosphosulfate (APS). APS can be further phosphorylated to form 3'-phosphoadenosine-5'-phosphosulfate (PAPS) which serves as sulfate donor in several reactions (Schmidt and Jäger, 1992).

However, most of the APS is converted to sulfite, performed by **APS reductase** (APR, formerly called APS sulfotransferase). The subsequent reduction of sulfite to sulfide is catalyzed by **sulfite reductase** (SIR).

ATP-S has been assumed as the rate limiting step enabling and initiating sulfur metabolism. Unexpectedly, molecular manipulation of ATP-S activity in tobacco (Hatzfeld et al., 1998) and Indian mustard (*Brassica juncacea*, Pilon-Smits et al., 1999) did not result in severe effects on plant growth and physiology indicating a minor role only of ATP-S in controlling sulfur metabolism in plants. Although there are no genetically engineered plants available yet for analysis, it can be assumed that APR is a prime regulation point in SO_4^{2-} assimilation (Brunold and Rennenberg, 1997). Its activity and steady-state mRNA level increase markedly and co-coordinately in response to sulfate starvation (Takahaski et al., 1997; Saito, 1999). Recently, it has been show in *Arabidopsis* that the expression of mRNA encoding APR was also induced by *O*-acetylserine (OAS, see below) (Koprivova et al., 2000). The last sulfate assimilation enzyme, SIR, is constitutively expressed (Bork et al., 1998).

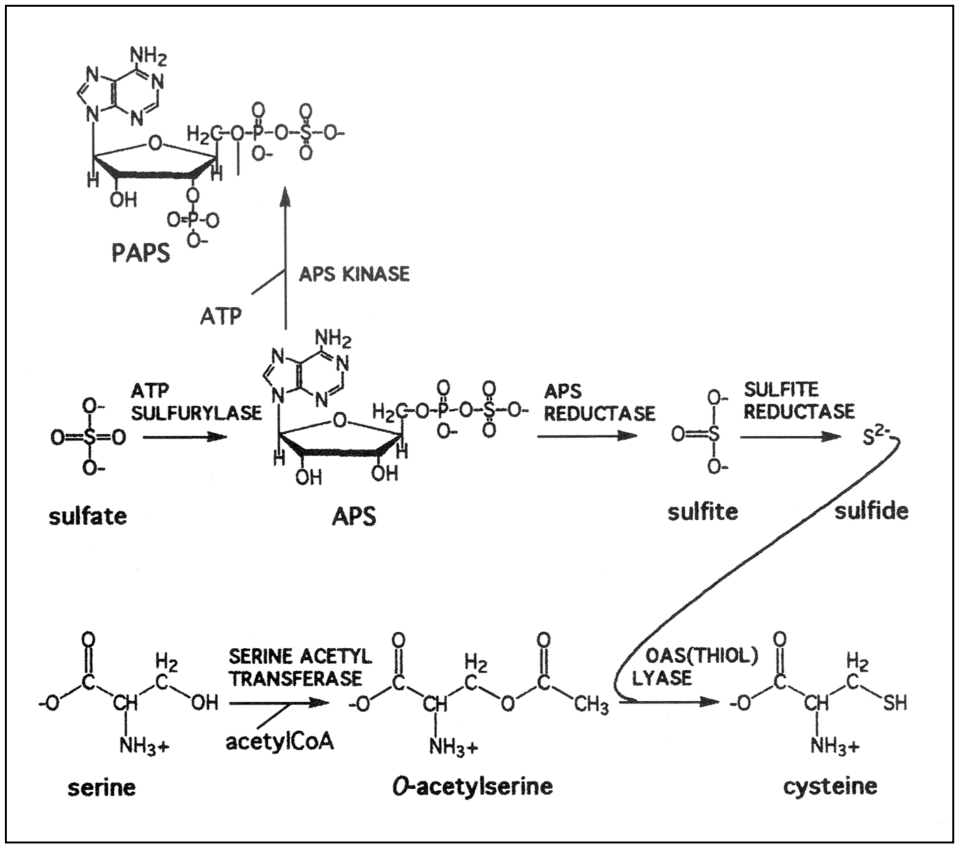


Figure 1.1: Cysteine synthesis requires coordination of the sulfate reduction pathway (top row) and the pathway leading to *O*-acetylserine synthesis (bottom row). Enzymes are indicated in upper case and metabolites in lower case letters (derived from Leustek et al., 2000).

The final step in cysteine synthesis (fig. 1.1) is the reaction that incorporates a sulfide moiety into the carbon skeleton derived from serine via *O*-acetylserine (OAS). A first enzyme needed is **serine acetyltransferase** (SAT) that is involved in the formation of *O*-acetylserine from serine and acetyl coenzyme A. This enzyme is at a branching point from serine metabolism to cysteine biosynthesis and consequently an important regulatory key enzyme. The activity of SAT is feedback inhibited by cysteine (Noji et al., 1998; Saito, 1999). However, the sensitivity of feedback regulation varies by enzyme sources and by isoforms, even from the same plants species (Noji et al., 1998).

The second enzyme ***O*-acetylserine thiol-lyase** (OAS-TL) is responsible for the terminal step, the formation of cysteine. cDNA clones encoding different isoforms of OAS-TL have been isolated from various plants (Grossman and Takahashi, 2001 and references therein). SAT is associated with OAS-TL through protein-protein interactions into a bi-enzyme complex, also known as cysteine synthase (Bogdanova and Hell, 1997; Droux et al., 1998).

An important feature of cysteine formation in plants is that SAT activity is 100- to 300-fold less in comparison to the activity of OAS-TL (Ruffet et al., 1994). In agreement with this observation, the availability of OAS seems to be the limiting factor for cysteine biosynthesis *in vitro* (Saito et al., 1994) and *in vivo* (Blaszczyk et al., 1999; Harms et al., 2000). Constitutive over-expression of a bacterial SAT gene in transgenic tobacco (Blaszczyk et al., 1999) and potato chloroplasts (Harms et al., 2000) resulted in increased levels of both cysteine and glutathione. This demonstrates the important role of SAT in the cysteine biosynthetic pathway under normal conditions, without any sulfur stress.

Overall regulation

SO_4^{2-} uptake, reduction and assimilation into cysteine is highly regulated in plants both at the level of gene expression and enzyme activity and further complicated by the subcellular localization of enzymes as well as developmental and spatial activity patterns and differences between plant species (Saito, 1999; Leustek and Saito, 1999). Induction of gene expression patterns under sulfur-deprived conditions has been systematically and thoroughly investigated in *Arabidopsis thaliana*. Under sulfur deficient conditions, expression of sulfur transporters and APR is predominantly induced in roots and expression of SAT and OAS-TL, though the latter to a lower extent, is induced in leaf tissues (Takahaski et al., 1997; Saito, 1999; Leustek and Saito, 1999). The signaling pathways are still under debate mainly discussing OAS as intracellular signal and glutathione and/or S-methylmethionine (SMM) as long distance transport molecules of reduced sulfur and as potential signal (Höfgen et al., 2001). Investigations of the effect of sulfur starvation on enzyme activity revealed that also the activities of sulfate transporters, APR, SAT and OAS-TL are increased. Metabolic control of sulfate uptake and further steps are shown to be mediated by several

factors: sulfate activation through ATP-S is reduced by cysteine and methionine, sulfate reduction through APR is inhibited by cysteine and methionine and stimulated by OAS and light, and conversion of serine into OAS through SAT is inhibited by several reduced organic sulfur compounds as cysteine, glutathione and methionine (Kopriva, 1999).

Localization

The enzymes SAT and OAS-TL are distributed in three major compartments of the plant cell, cytosol, chloroplasts and mitochondria, whereas the other enzymes of the sulfur assimilation pathway are found almost exclusively in chloroplasts (fig. 1.2). This resulted in the hypothesis that the chloroplasts in leaves are the major sites of sulfur assimilation (Hell, 1997; Leustek and Saito, 1999). However, recent experiments have shown that sulfur assimilation is not restricted to the leaves alone. Also in developing seeds, activity of several enzymes needed for S reduction was detected (see seed storage proteins).

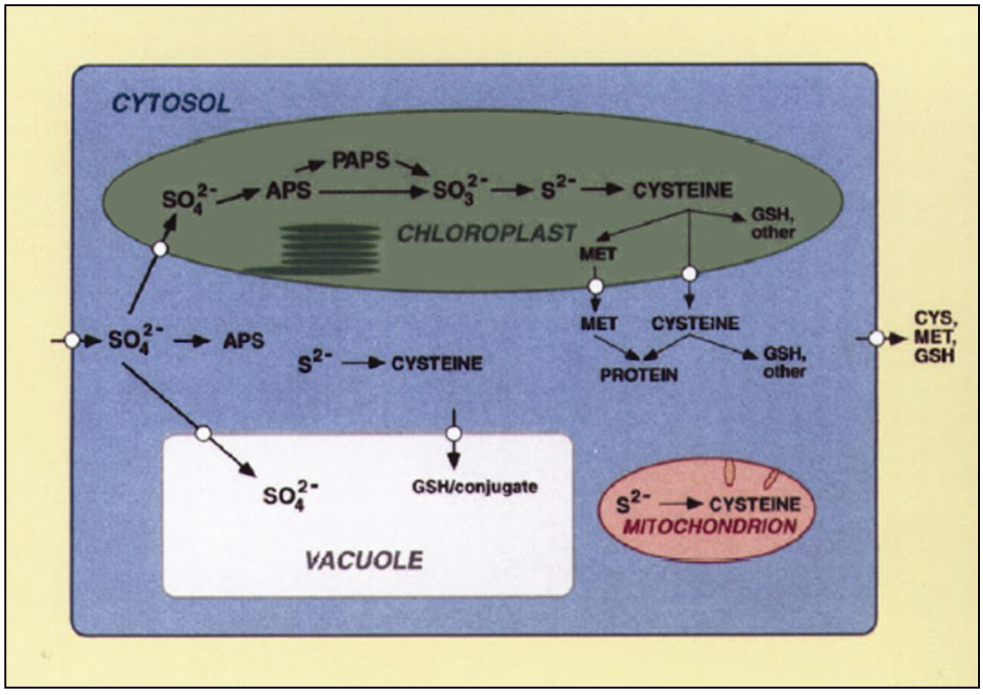


Figure 1.2: Subcellular compartmentation of major reactions and compounds of sulfur metabolism in a plant cell (taken from Hell, 1997).

Aspartate pathway

Pathway

In plants, methionine is synthesized from aspartate via several branches of the aspartate pathway along with lysine, threonine, and isoleucine (Azevedo et al., 1997; Matthews, 1999). A very similar pathway also operates in many bacterial species (Cohen and Saint-Girons, 1987).

The first step in the pathway (fig. 1.3) is the phosphorylation of aspartate by the enzyme **aspartate kinase** (AK). In plants, aspartate kinase has been studied for over 35 years and has been identified, isolated and characterized from a wide range of economical important crops like maize, barley, carrot, pea and soybean (Azevedo et al., 1997 and references therein). This enzyme has been found to be an important target for regulation, usually via end product feedback inhibition. There are multiple iso-enzymatic forms of the enzyme, which have different regulatory properties, including differences in their sensitivities to inhibition by lysine, threonine and S-adenosylmethionine (SAM), which is derived from methionine. Some forms are feedback inhibited by lysine alone, threonine alone or synergistically by lysine and SAM (table 1.1).

Plant *	Feedback inhibition #
Arabidopsis thaliana	LYS / THR
Barley	LYS / THR / LT / LM / LS / LSAM
Carrot	LYS / THR
Cucumber	LYS / THR
Lemna paucicostata	LYS / THR / LSAM
Maize	LYS / THR / LT / LTM / LSAM / SAM
Mustard	LYS / THR
Oat	LYS / THR
Pea	LYS / THR
Radish	LYS / THR
Rice	LYS / THR / LT
Rye	LYS / THR / LT
Sorghum	LYS / THR / LT
Soybean	LYS / THR
Spinach	LYS / THR
Sunflower	LYS / THR
Tobacco	LYS / LSAM
Wheat	LYS / THR / LT
Vinca rosea	LYS / THR

Table 1.1: aspartate kinase isolated from higher plants (taken from Azevedo et al., 1997).
(*) Feedback inhibition may vary according to the tissue of the plant tested. All different tissues tested for the plants listed were considered together.
(#) LYS = lysine, THR = threonine, SAM = S-adenosyl-methionine, LSAM = lysine plus SAM concerted inhibition, LT = lysine plus threonine concerted inhibition, LM = lysine plus methionine concerted inhibition, LTM = lysine plus threonine plus methionine concerted inhibition.

Aspartate semialdehyde dehydrogenase (ASDH) catalyzes the conversion of β-aspartyl phosphate to aspartate semialdehyde. This enzyme seems not to have any regulatory control of the pathway.

Aspartate semialdehyde can be converted to 2-3-dihydrodipicolinic acid by **dihydrodipicolinate synthase** (DHPS), the first enzyme committed to the synthesis of lysine. The activity of DHPS is competitively feedback-inhibited by lysine relative to aspartate semialdehyde and is the main regulatory step controlling lysine synthesis (Bryan, 1990). DHPS is approximately 10-fold more sensitive to lysine inhibition than are plant lysine-sensitive AKs (Galili, 1995).

Aspartate semialdehyde can also be converted to homoserine. This reaction is the starting point of the pathway branch leading to the synthesis of threonine and methionine. The enzyme catalyzing this conversion is **homoserine dehydrogenase** (HSDH) of which at least two isoforms have been observed in plants, one sensitive and the other insensitive to threonine inhibition (Bryan, 1980; Wallsgrove et al., 1983). It is possible that HSDH activity plays only a secondary role in limiting threonine synthesis in plants and that AK helps regulating threonine synthesis (Galili, 1995, Matthews, 1999). At least some aspartate kinase isoenzymes in plants appear to be bifunctional proteins and exhibit also homoserine dehydrogenase activity (Wilson et al., 1991; Azevedo et al., 1992). This observation was not unanticipated because two of the *Escherichia coli* AK isoforms exist as a similar bifunctional protein: AKI-HSDHI inhibited by threonine and AKII-HSDHII repressed by methionine (Cohen and Saint-Girons, 1987). In *Arabidopsis thaliana,* for example, there are at least two bifunctional isoforms (Ghislain et al., 1994), both threonine-inhibited and two monofunctional isoforms of AK (Tang et al., 1997; Frankard et al., 1997), inhibited by lysine and synergistically by S-adenosylmethionine.

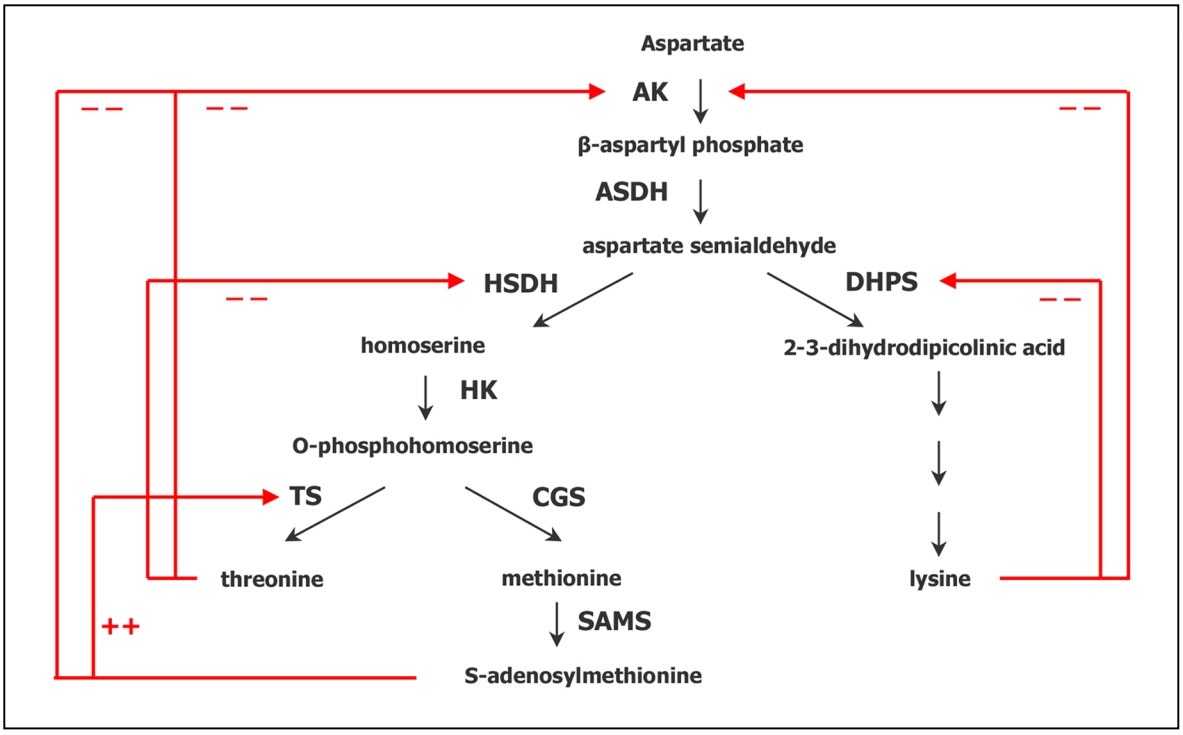


Figure 1.3: Overview of the most important steps in the aspartate pathway. The (--) symbols indicate feedback inhibition and (++) indicate stimulation of enzyme activity.

Once aspartate semialdehyde is formed, two enzymes compete for the substrate. This means that the regulation controlling enzyme activities of DHPS and HSDH discriminate whether aspartate semialdehyde will be channeled into the threonine-methionine synthesis branch or into the lysine branch. Experiments showed evidence that HSDH is a weak competitor for DHPS regarding their common substrate what indicates that lysine synthesis is favored over threonine and methionine synthesis. These experiments implied analyzing plants obtained by crossing a lysine-overproducing mutant with a threonine-overproducing mutant (Frankard et al., 1992) or by transforming with a bacterial feedback-insensitive DHPS and AK (Shaul and Galili, 1993). These crosses or double mutants resulted in plants overproducing higher levels of lysine, but never threonine showing that the flux of aspartate semialdehyde is almost entirely directed towards lysine synthesis.

Homoserine kinase (HK) catalyzes the reaction in which homoserine is converted to yield *O*-phosphohomoserine. The regulatory properties observed for this enzyme have been seen to differ considerably among plant species. For monocotyledonous plants, Riesmeier and co-workers (1993) suggested that HK is not a regulatory enzyme of the aspartate pathway. In contrast, HK from pea seedlings (Thoen et al., 1978) and radish leaf (Baum et al., 1983) was subject to inhibition by isoleucine and S-adenosylmethionine.

In plants, *O*-phosphohomoserine is the last common intermediate used to synthesize threonine and methionine. Therefore, the branch point enzyme **cystathionine-γ-synthase** (CGS) leading to methionine synthesis competes with **threonine synthase** (TS) leading to threonine synthesis for the pathway intermediate, *O*-phosphohomoserine (OPH). A striking parallelism can be found with the branching point DHPS-HSDH. In abundance of OPH, it is TS that funnels all excess of substrate into threonine synthesis. This is evidenced by the threonine-overproducing mutant (Frankard et al., 1991) and by transgenic plants expressing a bacterial feedback-insensitive AK (Shaul and Galili, 1992). Both plants overproduced threonine while there is almost no increase in methionine synthesis. In addition, a decrease in threonine synthesis due to a lower threonine synthase activity, leads to a significant increase in methionine (Bartlem et al., 2000; Zeh et al., 2001; see *de novo* synthesis of methionine).

Localization

Many of the enzymes involved in the aspartate family pathway have been localized to plastids (Bryan, 1990). Analyses of cloned DNA sequences confirm that these enzymes are synthesized with transit peptides that direct them into the chloroplasts (Galili, 1995). Operation of the aspartate pathway inside the plastid makes biological sense because some of the biochemical reactions require energy that is largely produced within this organelle.

Not only leaves harbor chloroplasts, also developing seeds contain appreciable levels of chlorophyll and have shown to be capable of photosynthetic electron transport to generate energy (Atkins and Flinn, 1978; King et al., 1998). This energy can also be used to make the aspartate pathway operational in seeds (see seed storage proteins).

Several of the key enzyme activities, AK, HSDH and DHPS are most abundant in rapidly growing tissues and in cell suspension cultures. Cellular proliferation and growth require protein synthesis, and hence amino acids. As plant tissue matures, the demand for protein synthesis diminishes and the activities of amino acid biosynthetic enzymes decrease. High activities of AK, HSDH and DHPS have been found in cell suspension cultures, young leaves and cotyledons (Matthews, 1999).

Overall regulation

The aspartate pathway converting aspartate into the derived amino acids lysine, threonine, methionine and isoleucine is mainly regulated by a negative feedback regulation of the end products on enzymes functional at key (branching) points of the pathway as described above and depictured in figure 1.3. However, the biological relevance of these control mechanisms remains unclear. Some enzymes have several isoforms and the role each isoform fulfils in the plant is not yet known. The observed results with an *Arabidopsis* DHPS loss-of-function mutant (Craciun et al., 2000), strongly suggest that the inhibition mechanisms might be more complicated *in vivo* than is supposed from *in vitro* experiments and that other, yet unknown, regulations are superposed upon feedback regulations of the key enzymes.

De novo synthesis of methionine

Pathway

The methionine molecule originates from three convergent pathways: the carbon backbone is derived from aspartate, the sulfur atom from cysteine, and the methyl group from the β-carbon of serine (Ravanel et al., 1998).

The carbon precursor of methionine synthesis in plants is different than in yeast and bacteria. In yeast, methionine is synthesized by direct sulfydration of *O*-acetylhomoserine while in bacteria succinyl-homoserine serves as substrate. Thus, in micro-organisms, homoserine is the branch point intermediate leading to the synthesis of methionine and threonine, whereas in plants *O*-phosphohomoserine is the last common intermediate to synthesize threonine and methionine (Datko et al., 1974).

In higher plants, starting from *O*-phosphohomoserine (OPH), methionine synthesis consists of two consecutive reactions, which constitute the trans-sulfuration pathway that consists in the transfer of the sulfur atom of cysteine (C3 skeleton) to homocysteine (C4 skeleton) with cystathionine as intermediate (fig. 1.4). First, **cystathionine γ-synthase** (CGS) catalyzes the condensation of cysteine and OPH to form cystathionine. cDNAs encoding CGS have been isolated from several plant species (Hesse et al., 2001 and references therein). **Cystathionine β-lyase** (CBL) subsequently carries out β-cleavage to produce homocysteine. CBL has been purified from spinach, *Arabidopsis thaliana* and *Echinocloa colunum* (Droux et al., 1995; Ravanel et al., 1996; Turner et al., 1998). Two isoforms could be distinguished of which one is located in the chloroplast whereas the other is cytosolic (Ravanel et al., 1998). The terminal step in methionine (Met) synthesis involves the transfer of the methylgroup from the β–carbon of serine to homocysteine by the enzyme **methionine synthase** (MS). So far, the molecular and biochemical characterization of MS from plants is still limited because of the small amount of protein which is present in plants (Hesse et al., 2001). This latter reaction is important not only in *de novo* methionine synthesis but also in the regeneration of the methyl group of S-adenosylmethionine (SAM). SAM is a key compound serving as a methyl donor in a variety of methyl transfer reactions, and it is the precursor of ethylene, biotine and certain polyamines (Ravanel et al., 1998). Biosynthesis of SAM is catalyzed by the enzyme **S-adenosylmethionine synthetase** (SAMS) and consists of adenylation of methionine. Recycling of the methyl group from SAM is performed with S-adenosylhomocysteine as intermediate.

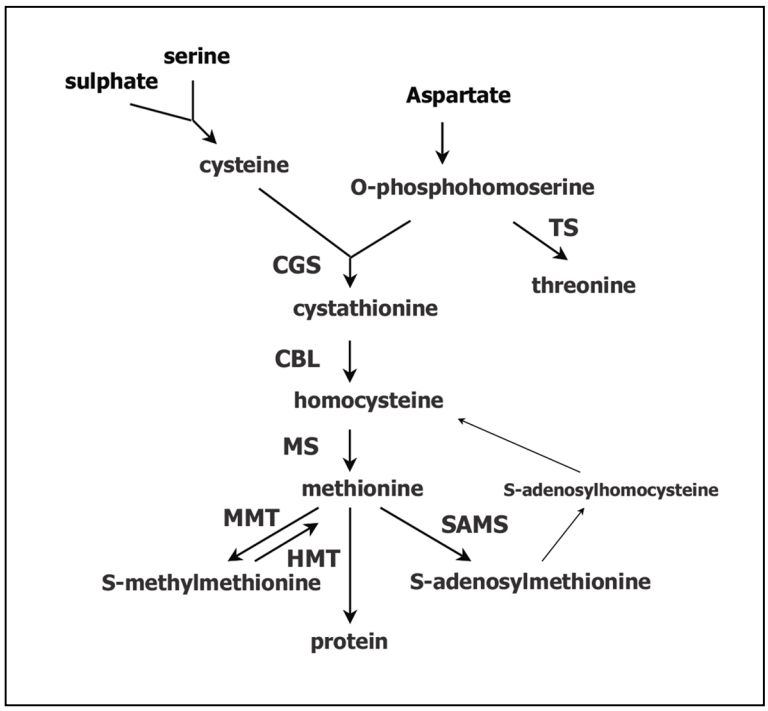


Figure 1.4: Methionine biosynthesis and metabolism in plants.

Methionine is also the precursor of S-methylmethionine (SMM). The postulated physiological role of this molecule is to serve as a storage reservoir and long distance transport form of labile methyl moieties and reduced sulfur as it is a major amino acid constituent in the phloem sap (Bourgis et al., 1999). SMM and Met are inter-converted through the action of **methionine S-methyltransferase** (MMT) and **homocysteine S-methyltransferase** (HMT) (Bourgis et al., 1999). Together these enzymes form the SMM cycle. Spatial separation of these enzymes such that SMM synthesis predominates in leaves and SMM-Met reconversion occurs in sink tissues, provides a mechanism for redistribution of Met (Hanson et al., 2000).

Localization

Most enzymes involved in Met production are localized to the chloroplasts in higher plants but the enzymes in the last part of the Met synthetic pathway make an exception. Cystathionine β-lyase (CBL), the second enzyme of *de novo* Met synthesis, is found in the chloroplast and cytosol (Droux et al., 1995) and methionine synthase, the last enzyme in Met production, has no apparent targeting peptide; thus, it is assumed to be cytosolic (Eichel et al., 1995). CBL activity measured in the cytosol seems to be non-specific (Ravanel et al., 1996; 1998) indicating that the conversion from cystathionine to homocysteine occurs exclusively in plastids. With respect to the localization of the final step, a new debate is arising, favoring the plastidial synthesis of methionine (Ravanel et al., 1998). The existence of additional methionine synthases however has still to be proven.

Next to the step from homocysteine to methionine, the conversion from Met to SAM and to SMM occurs in the cytosol as well (Trossat et al., 1996).

Overall regulation

Although it has been demonstrated that neither CGS nor CBL is regulated by aspartate-derived amino acids or pathway intermediates (Ravanel et al., 1996; 1998), the current understanding of the regulation of the methionine pathway remains limited.

According to the pioneering work carried out with *Lemna paucicostata* as a model plant, the key regulatory enzyme in the Met biosynthetic pathway is most probably cystathionine *γ*-synthase (CGS). A feeding experiment with *Lemna* showed that adding 2 *μ*M external Met decreased CGS-specific activity to 15 % of control, whereas supplementing with 36 *μ*M lysine and 4 *μ*M threonine to block Met synthesis, increased CGS activity two- to threefold (Thompson et al., 1982). Also, Giovanelli and colleagues (1985) determined that Met feedback regulates its own *de novo* synthesis at the step of cystathionine synthesis. It in now established that this negative regulation does not occur through feedback inhibition of CGS since this enzyme activity proved insensitive

to Met or derivatives (Ravanel et al., 1998). Rather, there is evidence suggesting that the gene expression of CGS is down regulated. Inaba and co-workers (1994) isolated an *Arabidopsis* mutant (*mto-1*) accumulating up to 40-fold more Met in a certain developmental stage. Analyses of CGS expression in these *mto-1* mutant plants revealed that the steady-state levels of CGS mRNA, protein and enzyme activity are three- to fivefold higher than in wild-type plants (Chiba et al., 1999). Application of Met to wild-type plants reduced the amount of mRNA for CGS, whereas no such effect was observed in the *mto-1* mutants. Further experiments showed that the *mto-1* mutants harbor a mutation in the CGS gene resulting in a more stable CGS transcript. The results suggested that an amino acid sequence encoded by the first exon of the wild-type form of the CGS gene acts in *cis* to down-regulate its own mRNA stability during translation and that this process is activated in response to excess Met. These findings demonstrate that CGS is post-transcriptionally regulated by Met or one of its metabolites and that unregulated expression of CGS leads to an increase in Met content. Saito (2000) mentioned that this amino acid stretch is highly conserved amongst several CGS sequences from different plant species and supposed that this regulatory mechanism would be functionally conserved in plant cells. Contradictory, an increased Met content in transgenic potato (as a result of antisense inhibition of threonine synthase, see below) had no detectable effect on mRNA or protein levels or on enzymatic activity of CGS (Zeh et al., 2001). This suggests that methionine synthesis is differently regulated between different species. Other experiments with *Arabidopsis* support the regulatory role of CGS in methionine biosynthesis. Constitutive over-expression of CGS in *Arabidopsis* plants led to increased accumulation of soluble Met and its metabolite SMM (Kim et al., 2002). Several transgenic lines showed silencing of CGS resulting in deformed plants with a reduced capacity for reproductive growth. Exogenous feeding of Met to the most severely affected plants partially restored their growth. Similar morphological deformities were observed in transgenic tobacco plants where SAM synthetase was co-suppressed (Boerjan et al., 1994), suggesting that these growth abnormalities are in part due to a reduced ability to produce SAM. These transformed tobacco plants accumulate up to 250-fold more free Met in their leaves than wild-type plants. It is clear, therefore, that methionine itself does not regulate the synthesis of any of the methionine-synthesizing enzymes. Consequently, SAM or some metabolite downstream of SAM, rather than methionine may be the negative regulator of CGS expression (Kim et al., 2002). Corresponding results were obtained by expression of a CGS antisense RNA and cDNA in *Arabidopsis thaliana* (Kim and Leustek, 2000; Gabière et al., 2000). Transgenic plants with up to 9-fold and 20-fold less CGS polypeptide and enzyme activity, respectively, revealed developmental abnormalities resulting in severe growth stunting and an inability to flower. Application of exogenous methionine to the transgenic plants reversed the morphological effects of CGS repression. However, only a slight change in the content of soluble methionine and its metabolite S-methylmethionine (SMM) was detected. This observation, together with other deviating results, points out that other regulation mechanisms are involved.

A first other important regulatory point in methionine biosynthesis is the competition for *O*-phosphohomoserine (OPH) between CGS and threonine synthase (TS) leading to the methionine and threonine branches of the aspartate pathway respectively. TS is strongly activated by low concentrations of SAM in a reversible and co-operative manner and it has a much higher affinity for OPH than does CGS. Thus, it has been proposed that CGS may compete poorly for OPH when Met (hence SAM) is abundant (Curien et al., 1996; 1998; Ravanel et al., 1998). By contrast, when Met (hence SAM) is limiting and TS activity declines, CGS has a greater ability to compete for OPH. Normally, the ratio of CGS to TS changes transiently and rapidly because of either increasing cellular concentrations of SAM or cysteine. Cysteine is responsible for the inhibition of TS (Giovanelli et al., 1984). Regulation of TS occurs at the level of enzyme activity rather than at the level of gene expression (Casazza et al., 2000). Transgenic tobacco plants constitutively expressing a bacterial TS enzyme yielded a 5-fold increase of free threonine (Muhitch, 1997), no effect on methionine level was observed, however. On the other hand, the *Arabidopsis* mutant *mto-2*, which displays a reduced endogenous TS activity, exhibits a 16-fold decrease in threonine content, accompanied by 22-fold elevated Met levels (Bartlem et al., 2000). The fact that methionine was able to over-accumulate when CGS levels were markedly reduced suggests that TS participates in regulating the methionine levels and that when TS activity is compromised, the regulation of CGS alone is not sufficient. Similarly, reduced TS activity by antisense inhibition resulted in decreased leaf threonine contents while methionine levels were increased tremendously by 60- to 239-fold in transgenic potato plants depending on the transgenic line and environmental conditions (Zeh et al., 2001). Remarkably (and not quite understood), tubers of TS antisense potato plants contain a Met level increased by a factor up to 30 and no reduction in threonine content. All these experiments indicate that the methionine biosynthesis is depending on TS activity by determining flux rates of OPH between threonine and methionine synthesis.

Another possibility is that cystathionine β-lyase plays an essential role. This was already suggested by the isolation of a methionine-mutant from protoplast cultures of *Nicotiana plumbaginifolia* by Negrutiu and colleagues (1985). The mutant showed a severe phenotype, stunted in growth and development. Supply of homocysteine and methionine in spraying experiments was able to restore growth to wild-type. In an experiment with transgenic potato plants expressing antisense CBL, Maimann and co-workers (2000) identified plants showing the same severe phenotype as the tobacco mutant of the group of Negrutiu (1985), supporting the data obtained for the mutant. Intriguingly, metabolites of the aspartate pathway and sulfur assimilation were also affected by this modification. Methionine decreased in concentration whereas cysteine, homoserine and cystathionine increased, demonstrating the reduced flow of Met precursors towards Met synthesis. These data show evidence for a regulation of CBL to modulate *de novo* synthesis of Met. Transgenic *Arabidopsis* plants expressing antisense CGS (Gabière et al., 2000) indeed displayed a two-fold increase in CBL level and enzyme activity over the control plants. Thus, it seems likely that antisense inhibition of CGS was compensated for by an increased CBL activity. However, further

experiments with transgenic potato plants constitutively expressing a CBL gene, showed no significant changes in content of amino acids and pathway intermediates when transgenic and wild-type plants were compared (Maimann et al., 2001). These data indicate that CBL is not in itself able to enhance metabolic flux towards methionine biosynthesis.

Overall regulation of the methionine pathway: summary

As can be seen in figure 1.5, the methionine pathway is very complex. This is because methionine receives its carbon skeleton from the aspartate family pathway, while its sulfur moiety is derived from cysteine. The synthesis of methionine is regulated at multiple levels. First of all, there is a general mechanism for control of the aspartate family amino acids centered on feedback inhibition of aspartate kinase (AK), the first enzyme of the aspartate pathway. The main part of the AK activity is tightly negatively regulated by lysine, in some cases in co-operation with S-adenosylmethionine (SAM), while threonine moderately regulates the other part of AK activity.

Secondly, methionine synthesis in plants is controlled at the level of competition between two key enzymes at branching points of the methionine pathway for their common substrate. This is the case with dihydrodipicolinate synthase (DHPS) and homoserine dehydrogenase (HSDH) competing for aspartate semialdehyde separating the lysine and threonine-methionine branch. HSDH is a weak competitor for DHPS what indicates that lysine synthesis is favored over threonine and methionine synthesis (see aspartate pathway). Also cystathionine γ-synthase (CGS) and threonine synthase (TS) compete for their common substrate *O*-phosphohomoserine (OPH) leading to the methionine or threonine branch of the pathway. Here, TS has a much higher affinity for OPH than CGS. The enzyme activity of TS is stimulated by SAM but inhibited by cysteine. Therefore, threonine synthesis is favored when methionine (hence SAM) is abundant. In contrast, when methionine is limiting, CGS has a greater ability to compete for OPH and methionine will be produced (see *de novo* synthesis of methionine).

Thirdly, *de novo* biosynthesis of methionine is mainly regulated at the cystathionine-forming step. In *Arabidopsis thaliana,* this occurs by post-transcriptional downregulation of the cystathionine γ-synthase (CGS) gene by SAM or some metabolite downstream of SAM (see *de novo* synthesis of methionine).

Fourthly, sulfur assimilation and cysteine synthesis are also highly regulated with APS reductase (APR) and serine acetyltransferase (SAT) as prime regulation points (see inorganic sulfur assimilation and cysteine synthesis). Not only is cysteine is needed as a precursor in Met biosynthesis but also the internal concentration of cysteine contributes to the regulation of Met biosynthesis. It inhibits threonine synthase (TS),

the branching enzyme of the threonine biosynthesis, and serine-acetyltransferase (SAT), the first enzyme in the biosynthesis of cysteine. Furthermore, it fulfils an inhibitory role in the activation and reduction pathway of sulfur together with the intermediate *O*-acetylserine (OAS).

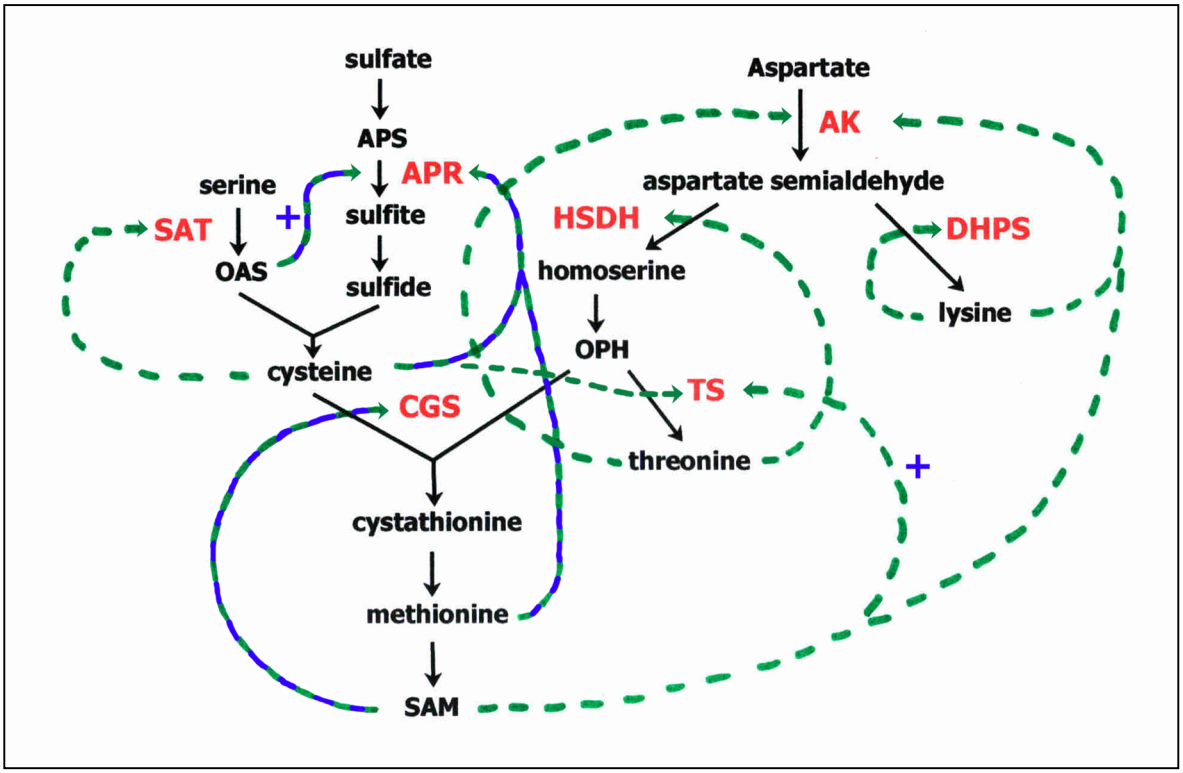


Figure 1.5: Overview of the methionine pathway. Only the key enzymes are mentioned: AK = aspartate kinase, DHPS= dihydrodipicolinate synthase, HSDH = homoserine dehydrogenase, TS = threonine synthase, CGS = cystathionine γ-synthase, APR = adenosine 5' phosphosulfate reductase, SAT = serine acetyl transferase. The most important regulatory effects are indicated with green arrows. The '+' symbol indicates a positive effect, in all other cases the effect is negative. Dotted green-blue arrows indicate effects on mRNA abundance, dotted green-white arrows indicate effects on enzyme activity.

In addition, recent experiments (Inaba et al., 1994; Gabière et al., 2000; Bartlem et al., 2000; Kim et al., 2002) suggest that the biosynthesis of Met is subject to complex developmental regulation. Also, it is clear that the tissue concentration of soluble Met is controlled by factors other than mentioned above. Thus, it is very likely that more parameters are involved in regulation of Met biosynthesis and further investigations are still needed to understand the complete regulatory circuit.

Moreover, the different seed amino acid balances, different sensitivity of key enzymes towards feedback inhibition, different concentrations of certain intermediates, … show that the methionine biosynthetic pathway is regulated in a different way in the different plant families.

Seed storage proteins

General features and classification of seed storage proteins

Mature seeds of most plants contain 8-25 % protein. This fraction includes proteins involved in seed development (i.e. metabolic proteins, housekeeping proteins, defense proteins and others), as well as storage proteins.

A seed storage protein may be defined as any protein accumulated in significant quantities in the developing seed that, on germination, is rapidly hydrolyzed to provide a source of reduced nitrogen for the early stages of seedling growth (Higgins, 1984). This comprises proteins that have no other known function than the one defined, as well as proteins that have additional biological activities. Consistent with their role as reserve of nitrogen for the germinating seedling, storage proteins tend to be rich in asparagine, glutamine, and arginine or proline while other amino acids are underrepresented. In legumes this generally is cysteine, methionine and tryptophan whereas in cereals, lysine, threonine and tryptophan are present at reduced levels. Other common properties of seed storage proteins are that they are all stored in the mature seed in the same subcellular compartments, the protein storage vacuoles or protein bodies, and that they show similar kinetics of accumulation and degradation.

Osborne (1924) classified seed storage proteins into groups on the basis of their solubility in water (albumins), dilute saline (globulins), alcohol/water mixtures (prolamins), and dilute acid or alkali (glutelins). Although there are exceptions, the major storage proteins of dicotyledonous plants are globulins and albumins and those of monocotyledonous plants are prolamins and glutelins. Sequence analysis of different seed proteins has revealed that the conventional solubility classes do not always consist of related proteins (Shewry and Tatham, 1990).

Seed storage proteins are encoded by several nonhomologous multigene families that vary in size, organization and chromosomal location (Casey, 1986; Kreis, 1985). The general features of seed storage protein gene expression programs are 1) expression is regulated temporally during embryogenesis with the highest seed storage protein accumulation during the mid-maturation stage, 2) seed protein genes are expressed exclusively during embryogenesis or are expressed at very low levels in mature plant organ systems and 3) expression is regulated spatially within embryonic organ systems (Goldberg, 1989; Thomas, 1993).

Legume seed storage proteins

Legume seeds contain three major classes of storage proteins. Taken together, these constitute more than 80 % of total seed protein and are therefore among the components that provide a major contribution to the nutritional quality of the seeds, as well as to the functional properties of products derived from them. Legume storage proteins were first described by Osborne as salt-soluble proteins (Osborne, 1924) and were grouped into two classes, vicilins and legumins, distinguishable in size and sugar content. Although both classes are present in most legumes, their relative abundance is highly variable, and some species are virtually devoid of either one or the other protein complex (Dudman and Millerd, 1975). A third protein class is represented by lectins, or carbohydrate-binding proteins, and the proteins related to them. These proteins are widely distributed in nature and, in trace amounts, may be found in most plant tissues. Although legume seeds normally contain lectins at relatively low levels, high amounts may be found in the tribe *Phaseolae*. In the case of seeds of some wild *Phaseolus vulgaris* accessions, the lectin-related protein arcelin is the most abundant seed protein, reaching levels up to about 40 % of total protein (Goossens et al., 1999).

Legumins are unglycosylated globulins with sedimentation coefficients in the range of 11-12S and have a complex hexameric structure. They are relatively rich in sulfur amino acids, compared to the other legume storage proteins. The constituent unit is made up of two disulfide-bonded subunits, which derive from posttranslational proteolysis of a single precursor polypeptide (Spencer and Higgins, 1980; Sengupta et al., 1981; Chrispeels et al., 1982b). Six of such pairs of subunits (an acidic α-subunit and a basic β-subunit) are non-covalently assembled into the 11-12S accumulated form of the legumin. Pea (*Pisum sativum*) legumin and soybean (*Glycine max*) glycinin can be taken as typical representatives of this protein class. Like most storage proteins, legumins are the product of a family of homologous genes, therefore both α- and β-subunits show a small degree of heterogeneity. It has been shown that in almost every legume species studied, legumin represents a major storage protein. However, this is not the case in *P. vulgaris*; even to that extent that only recently the first evidence for the presence of legumin proteins in *P. vulgaris* seeds has been reported (Mühling et al., 1997).

Seed proteins belonging to the so-called 7S class, or **vicilins**, are trimeric proteins that lack cysteine residues and hence cannot form disulfide bonds. Also the methionine content tends to be minimal, resulting in a very low amount of S-containing amino acids. Vicilins are usually less abundant than legumins, but there are remarkable exceptions; in *P. vulgaris* and *P. lunatus* vicilin is the major storage protein (Durante et al., 1989). Their detailed subunit compositions vary considerably, mainly because of differences in the extent of post-translational processing (proteolysis and glycosylation). In cultivated bean varieties, **phaseolin** accounts for up to about 60 % of total protein (Bollini and Chrispeels, 1978).

Legume **lectins** are a family of highly homologous proteins. Their peculiar characteristics are the presence of binding sites for Ca^{2+} or Mg^{2+}, required for the specific carbohydrate-binding and cell-agglutinating abilities of the protein. Several biological functions have been suggested for plant lectins. So far, roles in host specificity in *Rhizobium* nodulation and in defense against seed-eating insects have been proven (Diaz et al., 1989; Murdock et al., 1990). Among the different legumes, *P. vulgaris* seeds contain the highest amounts of lectins and lectin-related proteins. These are the carbohydrate-binding phytohemagglutinin (PHA) and the lectin-like proteins arcelin and α-amylase inhibitor. **PHA** may be considered the true bean lectin and normally represents up to 10 % of total seed protein (Bollini and Chrispeels, 1978). This protein is totally devoid of sulfur amino acids and is highly toxic to monogastric animals, therefore contributing to the lowering of the nutritional value of raw beans (Pusztai *et al.*, 1979). **Arcelin** is no or a very weak agglutinin and has been identified only in noncultivated wild bean accessions, where it may represent the major seed protein (Osborn et al., 1986). The bean **α-amylase inhibitor** is a heat stable glycoprotein active against mammalian and insect (but not plant) α-amylases and can account for up to 2 % of total seed protein (Powers and Whitaker, 1977).

Synthesis, assembly and deposition of seed storage proteins

Proteins accumulate in high amounts during the second stage of seed development, i.e. the mid-maturation stage (fig. 1.6). This stage is preceded by the zygote development stage and followed by the desiccation stage in which seeds undergo preparation for dormancy (Goldberg et al., 1989).

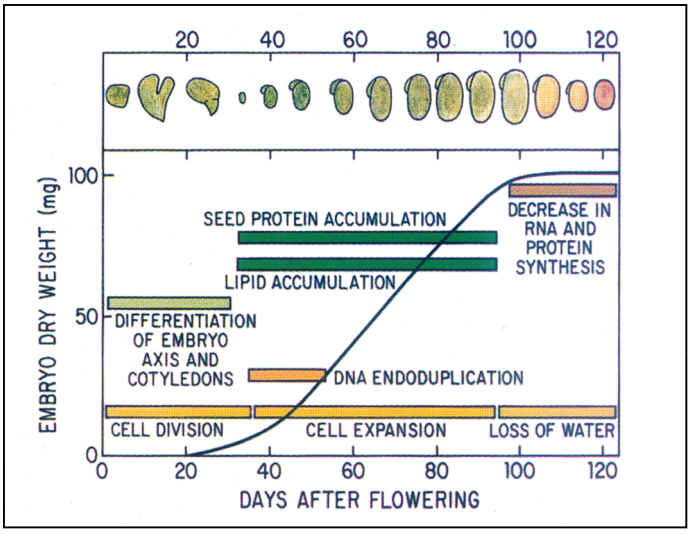


Figure 1.6: Major physiological events during embryogenesis (derived from Goldberg, 1989).

During the first phase following fertilization, cell division within the zygote gives rise to a globular embryo that differentiates into the embryonic axis and the cotyledons. In various dicotyledonous plants the second stage is characterized by significant mobilization from the endosperm to the embryo and the initiation of the synthesis of storage proteins in the embryo. At this stage, the embryo cells exhibit high activity of storage protein synthesis with no link to cell growth and proliferation. Only cell expansion takes place. This is followed by the third stage that consists of developmental arrest and desiccation, which prepares the seed for dormancy. The seed loses more than 90 % of its water content, RNA and protein synthesis terminate and embryonic dormancy begins. Although metabolic processes and *de novo* protein synthesis drops dramatically, desiccation-related genes are inducted during this stage. Dormancy ends in seed germination when seed storage proteins are used for the seedling development. The duration of different stages of seed development varies with the species and environmental factors (Goldberg et al., 1989).

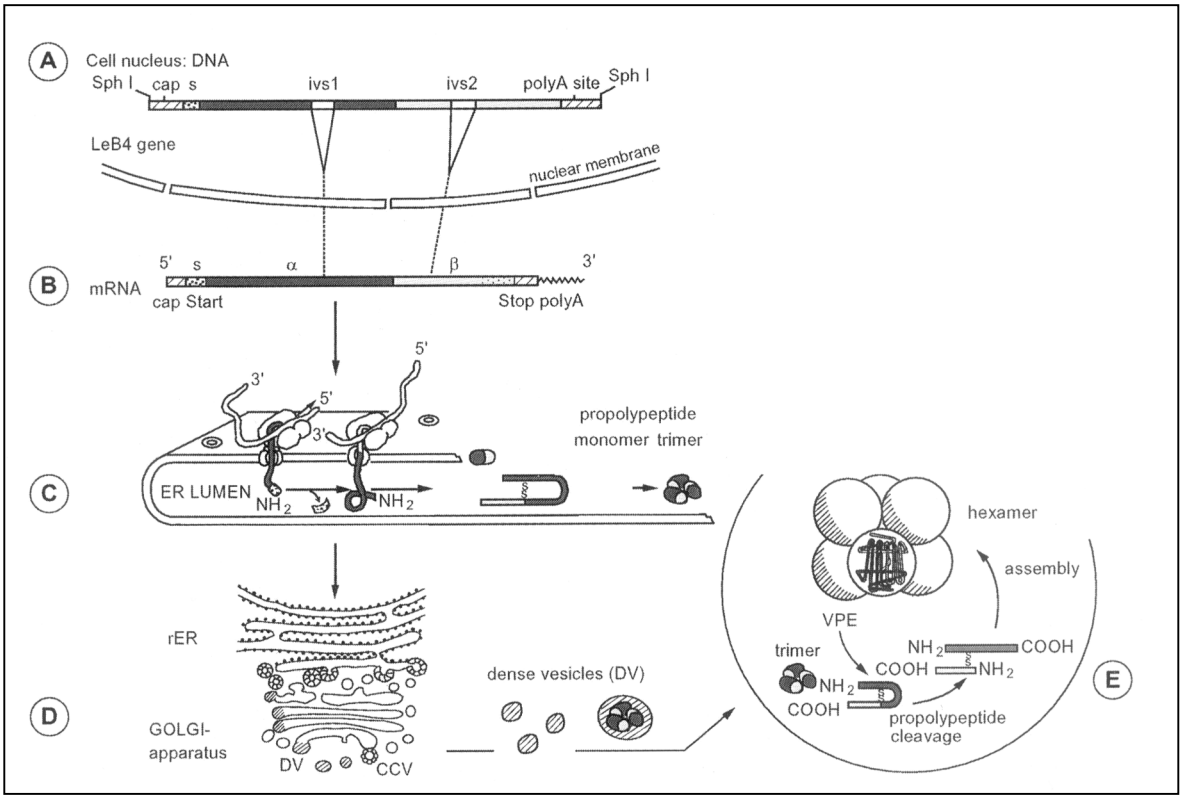


Figure 1.7: Diagram of legumin biosynthesis and sorting. In the cell nucleus (A) transcription generates the mRNA that specifies the signal peptide (s) and the α- and β-chain of a legumin subunit (B). The mRNA is translated on the rough ER (rER) where the nascent pre-prolegumin is segregated into the ER lumen with cotranslational cleavage of the signal peptide. After disulfide linkage formation prolegumin trimers are assembled in the ER lumen. They are transferred from the ER to the Golgi apparatus and sorted into dense vesicles (DV) at the trans-Golgi cisternae (D). Dense vesicles transport prolegumin trimers into the vacuolar compartment where prolegumin, in the trimers, is processed into α- and β-chains by the vacuolar processing enzyme (VPE) and trimer-to-hexamer transition takes place (E) (derived from Muntz, 1998).

Legume seed proteins, like all storage proteins, are initially synthesized on the **rough endoplasmatic reticulum** (ER, Chrispeels, 1991). This membrane system consists of an extensive, interconnected network of tubules and cisternae (Staehelin, 1997) and serves as the gateway of the secretory pathway. This pathway delivers proteins through the endomembrane system, via the Golgi apparatus, to the vacuoles where they form protein storage vacuoles (fig. 1.7). In the past, the terms protein bodies (PBs) and protein storage vacuoles (PSVs) have been used interchangeably, but PSV is now used to differentiate vacuoles containing storage proteins from PB originating from the ER (Galili and Herman, 1997).

The entry of storage proteins into the ER occurs cotranslationally and is specified by an N-terminal signal peptide that is cleaved from the nascent polypeptide chain as it enters the **lumen of the ER** (Von Heijn, 1984). After cleavage of the signal peptide, the precursors of storage proteins can be subject to core glycosylation and folding, processes facilitated by lumenal chaperones and enzymes (Boston et al., 1996; Vitale and Denecke, 1999). While still in the ER, disulfide bridges are formed which stabilize the structure and several polypeptides are joined to form an oligomer which has the proper configuration to be further transferred to the PSV. The major seed vacuolar storage proteins form dimers, trimers and tetramers in the ER lumen shortly after synthesis (Chrispeels et al., 1982a, 1982b; Ceriotti et al., 1995). Mutant proteins that are unable to form the correct quaternary structures are retained and degraded in the ER (Vitale and Denecke, 1999). This was, for example, the case for assembly-defective forms of phaseolin (Pedrazzini et al., 1997).

These oligomers are transferred to the **Golgi apparatus** where targeting to the vacuole appears to occur. Some storage protein precursors have short N- or C-terminal targeting sequences which are detached after arrival in the PSV. Others have been shown to possess internal sequence regions which could act as targeting information (Sanderfoot and Raikhel, 1999). Peptide targeting sequences have been identified for many storage vacuole constituents and there are several different types of unrelated sequences, but the targeting sequences for most of the major storage proteins, including the 11S and 7S globulin families as well as the seed lectins, have not yet been identified. The only identified targeting sequence is a C-terminal tetrapeptide of phaseolin which is sufficient for its vacuolar sorting (Frigerio et al., 1998; 2001). Another suggestion is that physical aggregation may serve as a sorting mechanism (Neuhaus and Rogers, 1998).

At the trans-Golgi cisternae, storage protein precursors are packaged into not well-characterized **dense vesicles** (DV) for transport to the PSVs. Such transport vesicles normally measure 0.1 to 0.2 µm in diameter (fig. 1.8) and contain storage proteins in a highly condensed state (Hohl et al., 1996).

Recent observations favor the hypothesis that there are different routes for seed storage proteins to exit the ER. Hara-Nishimura and co-workers (1998) proposed a route in which non-glycosylated storage proteins are directly delivered from the ER to

the vacuole, bypassing the Golgi complex. This pathway is mediated by **very large vesicles** (0.2 to 0.4 µm in diameter) that bud from the ER through an unknown mechanism and operates in parallel with the Golgi mediated pathway. These large vesicles are then incorporated into the PSVs, either by fusion or by autophagy. This alternative route was observed in developing seeds of pumpkin bean and pea (Robinson et al., 1995; Hara-Nishimura et al., 1998).

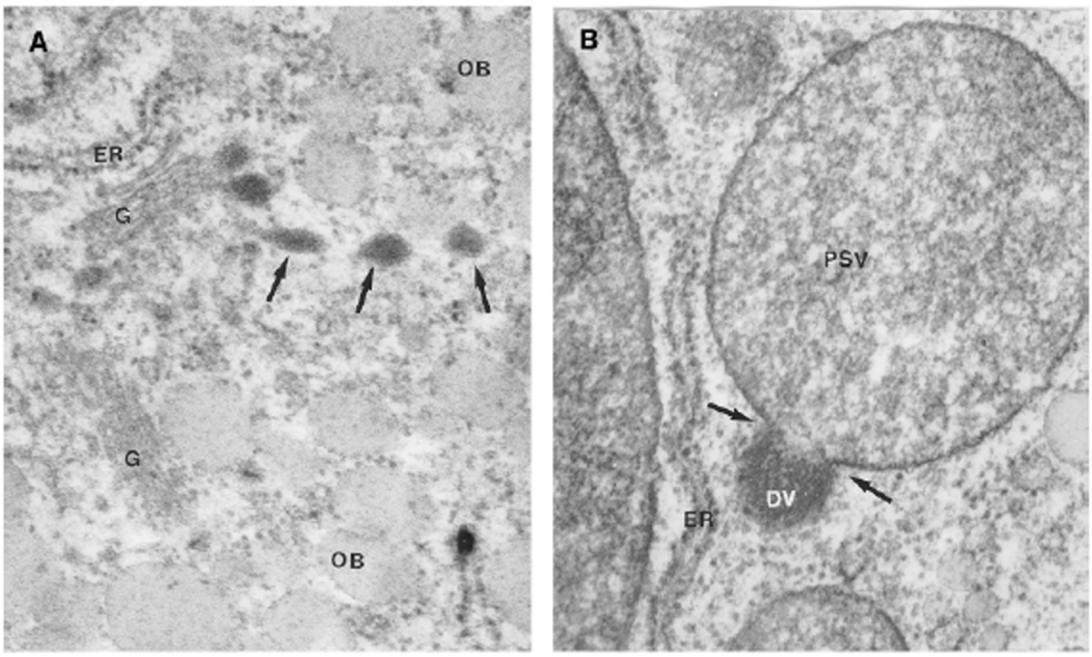


Figure 1.8**:** Ontogeny of organelles involved in protein storage in seeds. **(A)** Conventional electron microscopy of a mid-maturation soybean seed storage parenchyma cell showing the Golgi complex (G) secreting dense vesicles (arrows) that sequester storage protein precursors. OB, oil body. **(B)** The apparent fusion (arrows) of a dense vesicle (DV) to the protein storage vacuole (PSV) in a mid-maturation soybean parenchyma cell (taken from Herman and Larkins, 1999).

Protein storage vacuoles (PSVs) originate from post-Golgi central vacuoles that are devoid of significant accumulation in embryo cells. Storage proteins are added and gradually fill the vacuole. The accumulation of storage proteins within the vacuole is accompanied by additional processing that may serve to modify and prepare the proteins for dense packing. This processing includes modifications to both the polypeptide chain and glycan side chain, like polypeptide cleavage or ligation, removal of sugar residues from the glycan chains and further oligomerization. A nice example is the endoproteolytic cleavage of the 11S storage proteins into two chains linked by a disulfide bridge in seeds of monocots and dicots (Nielsen et al., 1995). The conserved cleavage site is on the C-terminal side of an asparagine, and the responsible asparagine-specific endopeptidase, termed vacuolar-processing enzyme (VPE), has been identified and its gene cloned (Hara-Nishimura et al., 1993, 1995). Nielsen and colleagues have demonstrated that VPE-mediated cleavage at the conserved asparagine site is required to convert the oligomeric trimer formed in the ER to the

mature vacuole-localized hexamer storage protein (Jung et al., 1998). Other seed proteins mature similarly through processing by VPE, although not all PSV proteins are modified after deposition. Within the *Phaseolus* storage proteins, phaseolin does not undergo fragmentation, although a few amino acids, probably from the C-terminal end, are removed (Bollini et al., 1982). Also phytohemagglutinin undergoes extensive proteolysis at the C-terminus (Young et al., 1995) as well as the bean α-amylase inhibitors, αAI1 and αAI2, and arcelin-5 (Young et al., 1999).

The vacuole not only serves as cellular protein storage compartment but as well as lytic compartment. PSVs contain numerous enzymes capable of completely degrading macromolecules, i.e. diverse acid hydrolases, including glycosidases, phosphatases, phospholipase D, and nucleases (Nishimura and Beevers, 1978; Mettler and Beevers, 1979; Chappell et al., 1980; Herman and Chrispeels, 1980; Van der Wilden et al., 1980). Moreover, recent studies suggest that PSVs contain active proteases even while storage proteins accumulate. But the conformation of the present seed storage proteins should provide resistance to co-localized PSV protease(s) (Herman and Larkins, 1999). An example here is the transition of a legumin trimer to a hexamer conformation that renders the legumin resistant against further attack by the vacuolar-processing enzyme (VPE) (Muntz, 1998).

Control of seed storage gene expression by sulfur-nutrition

Seed storage proteins are a reservoir of nitrogen and sulfur to be consumed during germination and plants will try to conserve as much nutrients as possible, considering what is available. Therefore, the seed storage protein composition will alter depending on the sulfur availability in the environment. Maternal plants that receive adequate S produce the full range of storage proteins in their developing seeds that is characteristic of the species. Those that receive inadequate S produce relatively little of the S-rich storage proteins and strongly accumulate S-poor proteins (Higgins, 1984; Wrigley et al., 1980).

As mentioned before, 7S globulins are much poorer in sulfur than 12S globulins, although both contain only small amounts of S-amino acid residues. As a consequence, legumin (12S) formation will be decreased and vicilin (7S) formation increased under conditions of S-deficiency, what results in a nearly constant globulin concentration if the degree of S-deficiency does not exceed certain limits. These results were reported from experiments with pea (Chandler et al., 1983; 1984; Randall et al., 1979), lupin (Blagrove et al., 1976; Tabe and Droux, 2002) and soybean (Holowach et al., 1984: Gayler and Sykes, 1985). In seeds of S-deficient pea plants a relatively high level of vicilin synthesis was observed throughout the entire phase of protein accumulation whereas legumin formation was reduced. In soybean seeds, the 7S and 12S globulins are called β-conglycinin and glycinin, respectively.

When the S-supply was insufficient, the accumulation level of glycinin was repressed whereas exogenous application of methionine enhanced the accumulation of glycinin. In another soybean report, the 12S/7S ratio was measured with different timing of sulfur deficiency (Sexton et al., 1998). This ratio was strongly influenced by S-deficiency occurring during reproductive growth. Provision of S near the middle of the seed filling period caused 12S/7S ratio to increase threefold over plants maintained S-deficient throughout seed filling.

Also experiments performed with the 7S globulin of soybean alone, demonstrated the influence of sulfur nutrition on seed storage protein accumulation. β-conglycinin is comprised of three subunits: a smaller β-chain which contains no methionine and two larger subunits, the α- and α'-chains containing low levels of methionine. Under S-deficiency, formation of the β-chains was increased and it was again decreased during recovery from S-deficiency. This has been investigated using cotyledon culture in nutrient solution with different levels of S-supply (Thompson and Madison, 1990) as well as using intact soybean plants (Gayler and Sykes, 1985). Also, when methionine was added to isolated cotyledons or fruit explants of soybean, the synthesis of the β-chain was suppressed (Holowach et al., 1984; Horta and Sodek, 1997).

If the corresponding genes were transferred into petunia or *Arabidopsis*, the pattern of regulatory dependence on S-supply was maintained in the transgenic seeds (Fujiwara et al., 1992; Naito et al., 1994). Even if the promoters of the β-conglycinin chains were fused to reporter genes, their expression in transgenic plants was similarly dependent on the degree of S-nutrition regardless of how sulfur was supplied, as sulfate or methionine (Hirai et al., 1994; 1995). Therefore, it appears that the S-dependent expression of genes for β-conglycinin chains is predominantly regulated at the transcriptional level. In contrast, the abundance of the S-rich protein pea albumin1 was restricted by destabilization of its mRNA in S-deficient pea seeds (Higgins et al., 1986). Thus, transcriptional and post-transcriptional mechanisms are implicated in decreasing accumulation of S-rich proteins in S-deficient conditions.

Regulation of other genes encoding S-rich proteins, like the proteinase inhibitors of soybean (Thompson and Madison, 1990), has also been reported to depend on sulfur supply.

Regulation of amino acids metabolism in plant seeds

The special structural characteristics of the seed storage proteins apparently require also a special amino acid composition for each storage protein. This issue is very relevant for the regulation of free amino acid metabolism in seeds because in order to be efficiently incorporated into storage proteins, the composition of free amino acids in plant seeds should match quite strongly the amino acid composition of the storage

proteins. Maintenance of such a delicate balance of free amino acids in the seeds is not simple because metabolism of amino acids in the seeds depends on transport of carbon-, nitrogen-, and sulfur-containing metabolites from the vegetative tissues. The level and ratios of the different transported metabolites are not fixed, but instead depend on complex, environmentally determined factors, such as photosynthesis as well as fertilization. Thus, seeds are expected to contain a supercomposite regulatory network of amino acid metabolism which can sense the arriving metabolites and convert them in the best way to the desired balance of free amino acids for most efficient incorporation into the storage proteins (Galili and Höfgen, 2002).

Synthesis of methionine in plant seeds

The major amino acids transported from the leaf canopy to developing seeds are asparagine and glutamine (Lea and Miflin, 1980). Other commonly transported amino acids include serine, alanine and glycine. Developing seeds are thus expected to possess the entire enzymatic machinery needed to convert these amino acids to all other amino acids necessary to produce seed proteins. On the other hand, S may be delivered to seeds in either a reduced or an oxidized form. A member of a family of sulfate transporters has recently been implicated in the process of S delivery to seeds in *A. thaliana* (Awazuhara et al., 2001).

Aspartate pathway

Experiments performed by Karchi and co-workers (1994), expressing a bacterial feedback insensitive AK and DHPS gene in a seed-specific manner resulted in a significant increase in free threonine and lysine respectively. These results indicate that the aspartate family pathway is functional in seeds. The timing of expression of the lysine and threonine biosynthesis genes as well as the metabolism of asparagine into aspartate and aspartate family amino acids seems to occur relatively late in seed development, similar to the timing of expression of seed storage genes (Karchi et al., 1994). Expression of an *Arabidopsis* AK/HSDH gene in transgenic tobacco seeds was also coordinated with the initiation and onset of storage proteins synthesis (Zhu-Shimoni et al., 1997).

S-assimilation and methionine biosynthesis pathway

It has been reported that soybean seeds could be cultured *in vitro* with sulfate as their sole sulfur source (Holowach et al., 1984). The fact that these seeds grew and accumulated storage proteins over a period of several days can be taken as indirect evidence for the ability of developing seeds to perform sulfur reduction and sulfur

amino acid biosynthesis. Sexton and colleagues (1998) tested the effect of timing of sulfur deficiency on seed storage protein expression in soybean. They concluded that the plant maintains a substantial capacity for S assimilation late into seed filling, and that mobilization of vegetative S is not a large source of S for developing seeds. Experiments performed by Tabe and Droux (2001) with developing lupin cotyledons, demonstrated that the sulfur arriving in the developing pods is predominantly in an oxidized form. They also demonstrated that developing lupin seeds were able to transfer the sulfur atom from ^{35}S-labeled sulfate into seed proteins *in vitro*, showing the ability of the developing cotyledons to perform all the steps of sulfur reduction and sulfur amino acid biosynthesis.

Sexton and Shibles (1999) detected significant quantities of ATP sulfurylase, the first enzyme of the pathway of sulfur reduction, in developing soybean seeds. Their experiments showed that seed tissue was a dominant site of ATP sulfurylase activity during seed filling. They believed that most of the S-amino acids synthesized during reproductive growth were synthesized in the seed itself, rather than being imported from other tissues. Two other enzymes of the sulfur amino pathway, serine acetyltransferase and *O*-acetylserine thiol-lyase were found in quantities theoretically sufficient to account for all of the sulfur amino acids that accumulate in the protein of mature lupin seeds (Tabe and Droux, 2001). In the same lupin seeds, cystathionine β-lyase activity was also detected providing the first proof for methionine biosynthesis in seeds. Presence of enzymes of the reductive sulfate assimilation pathway and methionine biosynthesis was also detected by Anderson and Fitzgerald (2001) in seeds of wheat. Extracts of endosperm of wheat supported ATP sulfurylase and *O*-acetylserine thiol-lyase activity, indicating that endosperm tissue has the potential to assimilate sulfate into cysteine. These extracts gave also evidence of cystathionine γ-synthase and cystathionine β-lyase activity, indicating that endosperm contains enzymes to perform methionine synthesis.

All these findings provide evidence that S-assimilation and biosynthesis of cysteine and methionine in the developing seed itself is an important source of sulfur amino acids for storage protein synthesis.

Arcelin-5 protein

Arcelins and other seed storage proteins of the common bean

Most cultivars of the common bean (*Phaseolus vulgaris* L.) contain two major seed storage proteins: the 7S globulin phaseolin and the bean lectin phytohemagglutinin. A number of electrophoretic variants of phaseolin are known with polypeptides in the 45 to 55 kDa range (Brown et al., 1981a), whereas phytohemagglutinin variants have polypeptides in the 33 to 41 kDa range (Brown et al., 1981b). In a few wild *P. vulgaris* accessions, a third important seed storage protein fraction is found: the lectin-like protein arcelin (Romero Andreas et al., 1986). Until now seven allelic arcelin variants have been identified. These variants have electrophoretic mobilities different from either phaseolin or phytohemagglutinin in the 30 to 40 kDa range (Osborn et al., 1986; 1988; Hartweck et al., 1991; Mirkov et al., 1994; Lioi and Bollini, 1989; Santino et al., 1991; Acosta Gallegos et al., 1998). The available sequence data show that the arcelin variants, phytohemagglutinin and the bean α-amylase inhibitor are all encoded by related members of a lectin gene family (Chrispeels and Raikhel, 1991). The genes of this family are tightly linked but unlinked to the genes encoding phaseolin (Osborn et al., 1986; Suzuki et al., 1995).

Many plant lectins probably have a defense function against insects and other predators (Chrispeels and Raikhel, 1991). Osborn and co-workers (1988) suggested that arcelins too would have such a role and proposed them to be responsible for resistance against the Mexican bean weevil (*Zabrotes subfasciatus*). Resistance towards this pest, an important cause of post-harvest losses in cultivated beans from tropical countries, was only found in wild *P. vulgaris* accessions containing one of the arcelin variants (Schoonhoven et al., 1983). However, when the insecticidal activity of the arcelin-5 variant that is present in the highly resistant *P. vulgaris* accession G02771 was investigated, no correlation could be established between the presence of the arcelin 5 and the insecticidal effects (Goossens et al., 2000). Insect feeding assays with artificial seeds, into which purified arcelin-5 protein was incorporated, and with transgenic *P. acutifolius* seeds, in which the arcelin-5 genes were expressed, showed that the presence of arcelin-5 proteins, even at elevated levels, was not sufficient to achieve adequate resistance against the Mexican bean weevil. Most probably, the resistance of the accession G02771 is due to the presence of another factor, genetically closely linked to the arcelin-5 allelle (Goossens et al., 2000). The putative arcelin-associated factor may act on its own or in concert with arcelin. An accessory or synergistic effect of arcelin might, for example, be related to its poor digestibility by larval enzymes of the Mexican bean weevil (Minney et al., 1990). In feeding tests with rats fed with beans from a *P. vulgaris* breeding line containing high levels of arcelin-1 no toxicological effects of the protein on mammals was found (Pusztai et al., 1993).

Characterization of the arcelin-5 genes

The arcelin-5 variant, found in the wild *P. vulgaris* genotype G02771, was characterized by Goossens and colleagues (1994, 1995). This genotype contains two arcelin-5 genes: the *arc5-I* gene that encodes the arcelin-5a protein (Arc5a, 32.2 kDa) and the *arc5-II* gene that encodes the arcelin-5b protein (Arc5b, 31.5 kDa) and a minor nonglycosylated isoform, the arcelin-5c protein (Arc5c, 30.8 kDa) (fig. 1.9). The sequence similarity between the *arc5-I* and *arc5-II* transcribed regions is more than 98%.

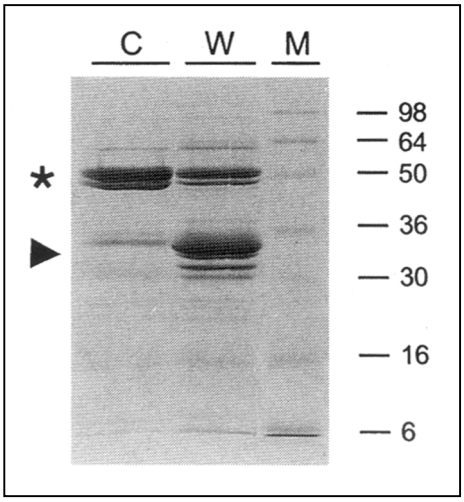


Figure 1.9: Crude seed protein extract of the wild *P. vulgaris* accession G02271 (W), which harbors the arcelin-5 genes and the cultivar Ica Pijao (C), which doesn't contain the arcelin-5 genes, separated by SDS/PAGE and visualized with Coomassie blue. Lane M contains marker proteins (molecular mass indicated on the left in kDa). asterisk = phaseolin, triangle = arcelin-5 proteins (derived from Goossens et al., 2000).

An 8.5 kb genomic clone harboring the *arc5-I* gene could be isolated (fig 1.10). The complete sequence of a 3.9 kb fragment was determined (Goossens et al., 1995). Analysis of the sequenced fragment of the *arc5-I* genomic clone (Goossens et al., 1999a) revealed the presence of a large number of putative regulatory elements in the 5' and 3' flanking sequences of the *arc5-I* gene. Among these are *cis*-regulatory elements thought to be involved in (quantitative) seed-specific expression (Thomas, 1993 and references therein). The analysis showed that most of the seed-specific motifs found in sequences of diverse legume globulins were also encountered in the 5' flanking sequence of the *arc5-I* gene. Aside from these putative seed-specific regulatory elements, computer analysis also showed the presence of multiple potential MARs (Breyne et al., 1994). MARs have attracted attention because of their perceived capacity to increase levels of transgene expression, reduce transformant-to-transformant variation of transgene expression, and confer copy number dependence to transgene expression. These properties are a consequence of a possible role for MARs as boundary elements or chromatin-regulatory elements (Holmes-Davis and Comai, 1998). Three clusters of potential MAR sequences were detected. Analogous situations were found in the regulatory sequences of other abundantly expressed seed storage proteins of leguminous species, such as the *P. vulgaris* phaseolin (Slightom et al., 1983) and the broad bean legumin B4 (Bäumlein et al., 1986).

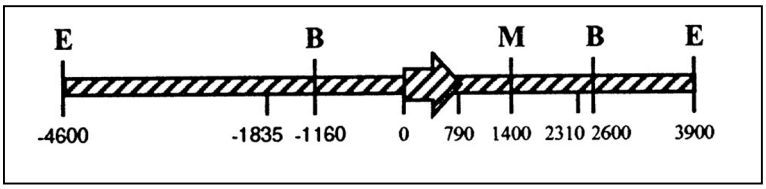


Figure 1.10: Schematic representation of the *arc5-I* genomic clone with indication of the coding region (arrow) and the restriction sites for *EcoRI* (E), *BamHI* (B) and *MunI* (M). Numbers correspond with positions in the *arc5-I* gene relative to the translation start site. Regions upstream of position –1835 and downstream of position +2310 were not sequenced.

In *Arabidopsis thaliana* transformation experiments (Goossens et al., 1999a), T-DNA constructs were used that contained *arc5-I* fragments of different sizes (fig. 1.10). The *Bam*HI/*Mun*I fragment contained potential TATA and CCAT boxes, the majority of the potential *cis*-regulatory elements for seed-specific expression and also one cluster of 5' MAR sequences. The second construct contained the *EcoRI* fragment, which represented the largest *arc5-I* fragment available from the genomic clone. This fragment harbored also the other potential MARs and possibly additional regulatory elements in the nonsequenced part of the *EcoRI* fragment. Although the maximum accumulation level obtained with both constructs was similar, the plants transformed with the construct containing the *EcoRI* fragment generally showed the highest expression levels (up to 15 % of the total seed protein). In both transgenic plant lines, a low plant-to-plant variation in transgene expression, among plants transformed with the same T-DNA construct, was observed. Moreover, no *arc5-I* transcripts were detected in total RNA preparations from flowers, cauline or rosette leaves, stems or roots from transgenic *Arabidopsis* plants, indicating that expression of the *arc5-I* gene was restricted to the seeds. These experiments indicate that the 5' and 3' flanking sequences as present in the isolated *arc5-I* genomic clone, contain all essential information for correct developmental and spatial regulation and exceptionally high accumulation of the arcelin-5a protein in transgenic plants.

Arcelin-5 is a very abundant protein (30-40 % of the total seed protein content), yet it is encoded by only two genes per haploid genome (Goossens et al., 1994). In contrast, phaseolin, which is the common seed storage protein present in all *P. vulgaris* genotypes, is encoded by a multigene family with seven to nine genes per haploid genome (Slightom et al., 1985). Phaseolin normally accounts for up to 60 % of the total protein of *P. vulgaris* seeds. It is not known whether all copies contribute equally to the observed expression levels, but this suggests that the amount of arcelin per gene copy is much higher than that of phaseolin.

Characterization of the arcelin-5 proteins

Primary sequences of legume lectins exhibit remarkable similarities, with a significant number of invariant amino acid residues. Most of these are involved in carbohydrate binding or in metal binding (Ca^{2+} and Mn^{2+}) which, in turn, is necessary for binding carbohydrates (Sharon and Lis, 1990). Sequence comparison showed that in arcelin-5 a number of these residues have been substituted while others have been deleted. Hemagglutination tests indicated that the arcelin-5 protein, indeed, has no or very low lectin activity (i.e. carbohydrate binding activity) (Goossens et al., 1994).

Both Arc5a and Arc5b are glycoproteins, but they have a different number of potential glycosylation sites (NXT or NXS) as could be determined from the deduced amino acids sequences of *arc5-I* and *arc5-II* (Goossens et al., 1994). Arc5a contains three potential N-glycosylation sites (Asn22, Asn70 and Asn79) (fig. 1.11), while Arc5b possesses only two sites (Asn70 and Asn79).

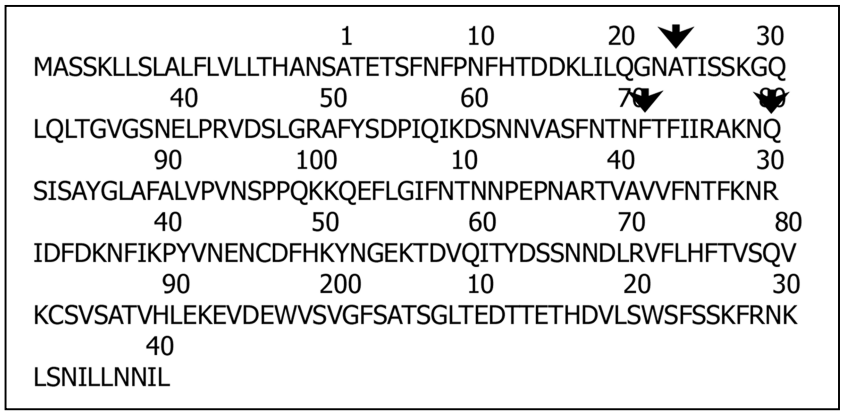


Figure 1.11: Amino acid sequence of the arcelin-5a protein (261 AA of which 21 AA code for the signal peptide and 240 AA for the mature Arc5a protein). The numbering on top starts from the N-terminus of the mature protein. Potential N-glycosylation sites are indicated with an arrow.

Electrospray mass spectrometry indicated that the Arc5a protein harbors two glycan chains and Arc5b one glycan chain (Goossens et al., 1994; Young et al., 1999). Examination of the crystallographic structure of Arc5a confirmed the presence of the glycan attached to Asn22. The other glycan was invisible, most probably due to its inherent flexibility (Hamelryck et al., 1996a). Glycosylation analysis showed that the glycan chains attached to Arc5a and Arc5b are most probably of the modified, fucosylated type (Goossens et al., 1994).

Of the three arcelin-5 proteins present in seeds of the *P. vulgaris* genotype G02771, Arc5a has the highest accumulation level (60 % Arc5a, 30 % Arc5b and 10 % Arc5c of total extractable arcelin-5 protein; Alain Goossens, personal communication). The arcelin-5a protein consists of 261 amino acids, of which the first 21 amino acids form the signal peptide (fig. 1.11). None of the amino acids of the mature protein is a methionine. Arc5a can exist either as a monomer or a dimer (Goossens, 1998).

The crystallographic structure of Arc5a was solved at a resolution of 2.7 Å (Hamelryck et al., 1996a). The overall structure is similar to the structures of the other solved legume lectin structures and consists of a flat six-strand β-sheet, called the back sheet, packed against a curved seven-strand β-sheet, called the front sheet (fig. 1.12). The structure contains 228 of the 240 amino acids of the mature protein, no electron density was observed for the terminal 12 amino acids. Young and colleagues (1999) found that the arcelin-5 protein undergoes C-terminal cleavage of 8 to 11 amino acids. The Arc5a monomer contains one disulphide bridge, between residues 146 and 182. These two residues link the sixth strand of the flat back sheet and the seventh strand of the curved front sheet.

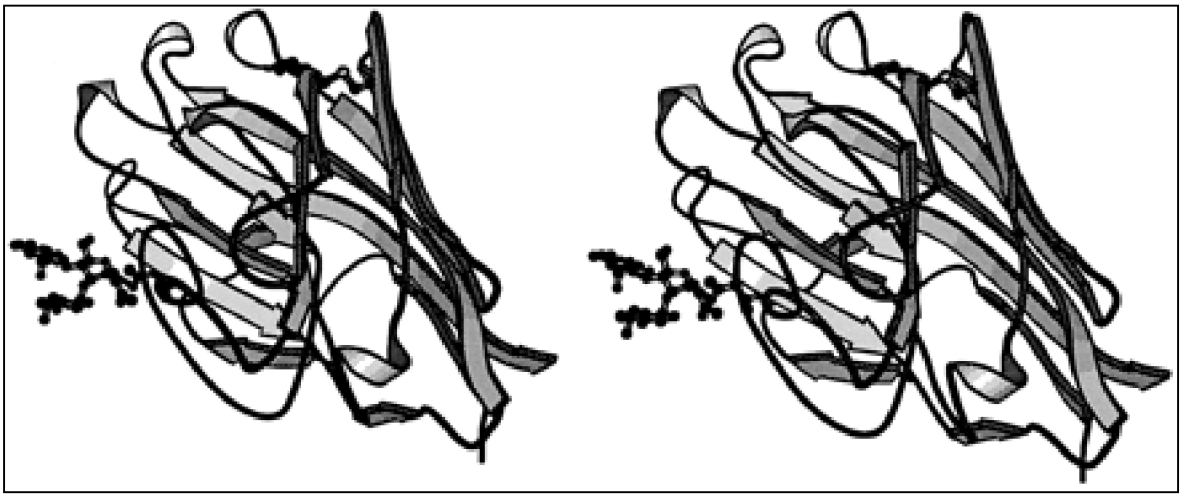


Figure 1.12: An overall view of the arcelin-5a monomer. The glucan attached to Asn22 and the disulphide bridge between residue 146 and 182 are shown as ball-and-stick model. This cysteine-bridge links the seventh β-strand of the front sheet and the sixth strand of the back sheet together (taken from Hamelryck et al., 1996a).

Strategies to improve the methionine content of legume seeds

Introduction

The seeds of plants are a superior food source because they are rich in carbohydrate, protein, and/or lipid, and unlike other plant organs, they possess a minimal amount of lignin and cellulose. Because humans and livestock are unable to synthesize all amino acids, the essential amino acids are required in their diets. No single seed type contains a complete regime of these essential amino acids. The proteins in cereal seeds are deficient in lysine and tryptophan, whereas legume seeds contain proteins deficient in the sulfur-containing amino acids, methionine and cysteine. Consequently, diets based on a single cereal or legume species result in amino acid deficiencies. This problem is reduced when a mixed diet of legume and cereal seeds is consumed, but the ability to produce these seeds is restricted in certain regions of the world. Eating is not only important for health, but is also considered one of the pleasures in life, and in many cultures, food has both social and religious value. Thus, trying to persuade people to change their traditional diet for nutritional reasons is difficult. For these reasons, it can be advantageous to modify the existing proteins of seeds to improve the composition of essential amino acids (Nelson, 1969). In highly developed countries, the need for a single seed source with a well-balanced amino acid profile is less critical, because diets are generally complex and include animal protein. Nevertheless, the use of seed proteins for livestock feed necessitates that an animal's diet meets a prescribed amino acid composition to promote efficient growth and produce meat. The conversion factor of feed protein nitrogen to body protein nitrogen rises with a better-balanced amino acid composition of the feed protein. Concomitantly, the amount of excreted nitrogen decreases as the conversion factor increases (Kirchgeβner et al., 1994). Thus, improving the amino acid composition of plant seed proteins can result in a decrease of the environmental pollution by nitrogen.

Methionine supplementation

A possibility to correct the methionine deficiency of seeds is to supplement it with its synthetic counterpart. Bressani and Elias (1968) demonstrated that supplementation of legume seeds with crystalline methionine could improve the growth of animals and the efficiency in their use of dietary protein. In practice, chemically produced methionine (fig. 1.13) is mostly used to supplement in the formulations of poultry and

pig feed. Except for soybean protein-based infant formulas, there is currently little use of free methionine to supplement human diets.

In the United States, methionine supplements to livestock feed amount to 50.000 ton each year and an annual cost of 120 million dollars to the growers (Beach and Ballo, 1992). These costs are direct expenses for farmers and therefore passed on to the consumers. Another disadvantage of using free methionine in supplemented foods and feeds is that it can be lost through leaching during processing, which can lead to formation of undesirable volatile sulfides through bacterial fermentation that generate off-flavors and odors. Acceptability is perhaps the biggest problem in supplementation of human diets with crystalline methionine.

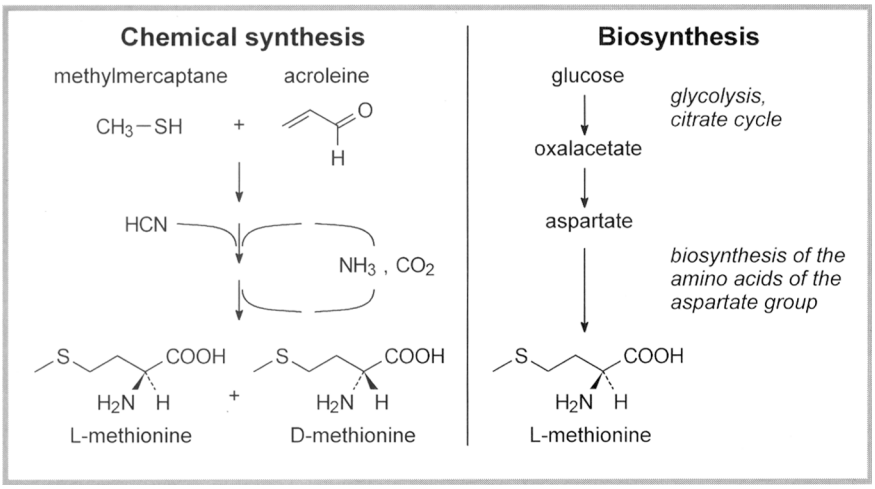


Figure 1.13: Chemical synthesis of methionine. As in every chemical synthesis of optically active molecules, 50 % D- and 50 % L-methionine is produced. Although living organisms utilize only L-form amino acids, both forms can be used as feed supplement by livestock such as poultry and pigs. These animals produce specific enzymes that can convert the D-form into the nutritionally active L-form (derived from Kircher and Huthmacher, 1999).

Breeding

Ever since the scientific basis for the nutritional limitation of seed proteins was established (Osborne, 1924), researchers have attempted to alter the amino acid composition of seeds. Until recently, this work was largely based on conventional breeding programs with a view to identify mutants that have elevated levels of the missing amino acids.

Mutants with a modified ratio among the different seed protein fractions

One strategy was searching for mutants where the seed protein fractions poor in the major limiting amino acids are decreased in favor of other protein fractions with a higher concentration of the limiting amino acids (Bright and Shrewy, 1983). Using this strategy, high-lysine mutants could be identified in cereals with reduced prolamin content, low in lysine, and a simultaneous increase in other lysine-richer protein fractions. Examples are opaque-2 and floury-2 in maize (Mertz et al., 1964), RisØ 1508 and Hiproly in barley (Shrewry et al., 1980) and IS-11167 and P-721-N in sorghum (Singh and Axtell, 1973). However, when seed collections of legumes and their near relatives, from a wide range of origins, were screened for lines that had a higher than average level of the sulfur-containing amino acids, it was clear that there was very little genetic variation for this character. For example, in one survey of 15 lines of pea and related species (Schroeder, 1982) any increase in one relatively sulfur-rich seed fraction (such as legumin) was invariably accompanied by a decrease in another major sulfur-rich fraction (the albumins). The consequence of this is that the S-amino acid content of total seed proteins remains fairly constant. The major seed-storage proteins of grain legumes vary in their content of methionine and cysteine. However, most seed storage proteins contain few or no sulfur-containing amino acids, resulting in a low overall seed protein sulfur-amino acid concentration. In common bean, phaseolin appears to contain the most sulfur amino acids among the major seed proteins. Gepts and Bliss (1984) selected for inbred backcross breeding lines with increased phaseolin per seed and found that these seeds had higher levels of available methionine indeed. But the increased methionine concentration was by far insufficient to produce greatly improved protein nutritional quality (Bliss, 1990). This successful strategy for cereals seemed not useful to increase the methionine content of legume seeds.

Mutants with a modified biosynthetic pathway

Another breeding strategy was producing and selecting plant mutants with an elevated level of synthesis and accumulation in the free amino acid pool due to a mutant enzyme in the biosynthetic pathway of the limiting amino acid. In all experiments performed, seeds were not yet the object of study.

Most of the investigations have been performed with enzymes of the aspartate biosynthetic pathway. The mutants were isolated by selection for growth in the presence of the lysine analogue, S-2-(aminoethyl)-1-cysteine (AEC) or high concentrations of lysine plus threonine (L+T). With AEC, lysine-overproducing plant mutants could be isolated for several species (for an overview see Galili and Larkins, 1999). The mutant plant possesses a larger pool of lysine what makes him resistant by being able to dilute out the toxic analogue. In tobacco (Negrutiu et al., 1984), lysine overproduction was due to expression of a mutant DHPS gene insensitive to the normal feedback inhibition by lysine (fig. 1.14). In the presence of L+T, the activity of

all AK isozymes is completely inhibited, resulting in methionine starvation (Jacobs et al., 1987). In this way, plants containing a mutant feedback-insensitive AK gene could be identified who overproduce free threonine and, to a lesser extent, methionine. This was the case for maize (Hibberd et al., 1980; Diedrick et al., 1990; Dotson et al., 1990), barley (Bright et al., 1982), carrot (Cattoir-Reynaerts et al., 1983), tobacco (Frankard et al., 1991) and *Arabidopsis* (Heremans et al., 1995; 1997).

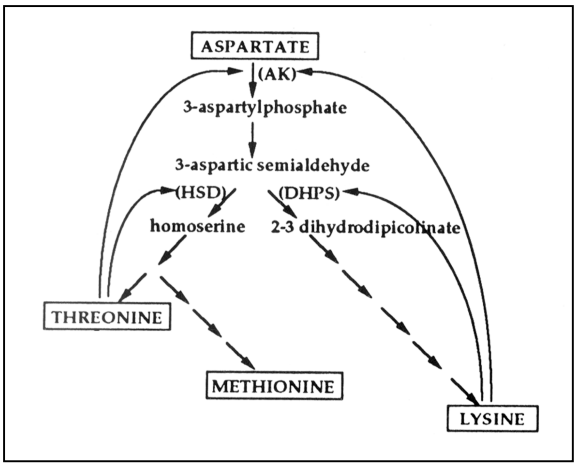


Figure 1.14: Diagram of the aspartate pathway. Only the major key enzymes and their products are indicated. Curved arrows represent feedback inhibition. AK = aspartate kinase, HSD = homoserine dehydrogenase, DHPS = dihydrodipicolinate synthase (adapted form Karchi et al., 1993).

To isolate plant mutants that over-accumulate methionine, the toxic analogue of methionine, ethionine, was used. Because of their similarity in chemical structure and composition, an increased concentration of methionine lessens the toxicity of low concentrations of ethionine. Soybean tissue culture cells selected for resistance to ethionine, contained up to 22 times higher levels of free methionine than controls (Madison and Thompson, 1988). These ethionine-resistant lines contained also an elevated internal concentration of S-methylmethionine (SMM) that serves as a methionine reserve. In *Arabidopsis*, the mutant *mto-1* was isolated that accumulates soluble methionine up to 40-fold in young rosettes (Inaba et al., 1994). As described before, this is the result of a mutation in cystathionine-γ-synthase (CGS), the first enzyme in *de novo* synthesis of methionine. In wild-type *Arabidopsis*, there is a mechanism to down-regulate the accumulation of CGS mRNA in response to excess of methionine and this mechanism is impaired in the *mto-1* mutant (Chiba et al., 1999). Recently, another *Arabidopsis* mutant, *mto-2*, was characterized that overproduces soluble methionine 22-fold in young rosettes (Bartlem et al., 2000). Here, over-accumulation of methionine was due to a mutation in the gene encoding threonine synthase (TS) that destroys the ability to synthesize threonine, resulting in channeling the shared intermediate OPH to methionine biosynthesis and overproduction of Met (see *de novo* synthesis of methionine).

In soybean, a mutant line was isolated using selection for ethionine resistance and visual phenotypic screening for methionine overproduction (Imsande, 2001). The seed methionine and cysteine concentrations of this line are each approximately 20 % larger than those of the parental line. The underlying causes are not yet investigated.

Most of these mutants are very useful to study the biosynthetic pathways of the different amino acids but are of no use to bring into practice because the mutant enzymes are expressed in all plant tissues, resulting in a constitutive amino acid overproduction. This overproduction may be toxic resulting in phenotypic alteration and sterility (Frankard et al., 1992). Other negative pleiotropic effects and yield depression have been detected in these mutant plants too (Bright and Shrewy, 1983; Glover et al., 1987).

Engineering the amino acid metabolism of the seed

Advances in plant tissue culture techniques and gene transfer technology have opened up new possibilities for modifying the amino acid content of plants. Recombinant DNA technology and plant transformation made it possible to express a gene coding for a mutant enzyme of the biosynthetic pathway of the limiting amino acid in the seeds of a chosen plant.

To increase the level of free methionine in legume seeds, the biosynthetic pathway leading to methionine (see before) can be engineered. Therefore, important enzymes at the beginning or at branch points of the aspartate pathway and *de novo* methionine synthesis, that are submitted to feedback inhibition, are targets to substitute with a mutated isozyme insensitive to feedback regulation. Mutant isozymes can be isolated from plants subjected to mutagenesis as described above, but also from bacteria since the methionine biosynthetic pathway is almost identical in micro-organisms as in plants.

Karchi and co-workers (1993) achieved the first successful enhancement of seed's methionine content. They isolated a mutant AK gene (*lysC*) from *Escherichia coli* that was desensitized to feedback inhibition by lysine and threonine (Shaul and Galili, 1992) and expressed this mutant gene in tobacco in a seed-specific manner. This expression was accompanied by a significant increase in the levels of free threonine (17-fold) and free methionine (3-fold) in the seeds of transgenic tobacco, but there was only a slight increase in the level of protein-bound methionine (6.8 %; Falco et al., 1995). The same insensitive AK gene was also expressed in the seeds of *Vicia narbonensis* (Pickardt et al., 1998) and the results were in accordance with these obtained with transgenic tobacco. The transgenic *Vicia* seeds accumulated three- to sevenfold more threonine and twofold more methionine than wild-type seeds.

Influencing the accumulation level of seed proteins by using the anti-sense approach

The overall amino acid profile of seeds is often determined by the combined amino acid composition of a few abundant proteins. It is possible that a minor protein may contain a higher content of methionine, yet contributes little to the final concentration of this essential amino acid because of a very low expression rate. Through genetic engineering technology, altering the ratios among the different seed protein fractions can be executed more specific and profound. The antisense method makes it possible to inhibit the expression of the gene encoding the dominant seed storage protein with little methionine content in favor of the methionine-richer protein. This approach was implemented successfully in *Brassica napus* (Kohno-Murase et al., 1995). Cruciferin is a 12S seed storage protein that accounts for 60 % of total seed protein of canola and it contains lower levels of methionine and cysteine than the second most prominent seed storage protein, napin (Simon et al., 1985). Seed-specific expression of an antisense gene for cruciferin reduced the cruciferin content and the reduction was balanced by an increase in napin content. This change resulted in increases of 8 % methionine, 32 % cysteine and 10 % lysine over the respective levels in non-transgenic seeds without changing the total protein content of the transgenic seeds significantly from that of normal seeds.

Most probably, this strategy is not useful to increase the methionine content of legume seeds because most legume seed storage proteins contain few sulfur-containing amino acids. One known high-methionine legume protein is a 10.8 kDa protein of soybean (Kho and de Lumen, 1988). This protein contains 12.1 % methionine but accumulates at a very low level, around 0.6 % of the total seed protein (George and de Lumen, 1991). Till now, no further experiments are published with this protein or any other legume methionine-rich protein.

Transfer and seed-specific expression of a heterologous gene coding for a high-methionine protein

In this strategy, a gene encoding a methionine-rich protein originating from any source organism is targeted for transfer into seeds of a selected plant. The expression of the transgene introduces a methionine-rich protein into the protein pool of the host, resulting in an increase in the methionine concentration of the seeds (table 1.2).

2S albumin of Brazil nut

The favored object for such experiments is the gene coding for the 2S albumin of Brazil nut, *Bertholletia excelsa* (BNA) that contains 18.8 % methionine and 7.9 % cysteine (Altenbach et al., 1987). In a first experiment, this gene was expressed under the seed-specific promoter of phaseolin in tobacco plants (Altenbach et al., 1989). Integration of different copy numbers of the chimeric gene was detected in maturing seeds and western analysis revealed that the methionine-rich protein was synthesized and stably accumulated in transgenic tobacco seeds. The level of the inserted protein varied with transgenic plant, ranging from 3 to 8 % of total seed protein, giving a methionine content that was enhanced by 10-30 % over normal seeds. Although BNA is also relatively rich in cysteine, no concomitant increase in the cysteine content was observed in the transgenic seeds. Since the content of the 2S albumin fraction and the total seed protein in the transgenic seeds remained the same as in normal seeds, it is assumed that the synthesis of some cysteine-rich 2S seed proteins was suppressed as a result of the introduced protein.

The same chimeric gene was introduced in *Brassica napus* plants (Altenbach et al., 1992). In the transgenic canola seeds, the heterologous methionine-rich protein accumulated at levels ranging from 1.7 to 4.0 % of the total seed protein, resulting in up to 33 % more methionine (fig. 1.15). In addition, Guerche and colleagues (1990) transformed canola plants with the gene encoding BNA, but using the soybean lectin promoter. Results revealed that the expression was tissue specific and developmentally regulated. The protein accumulation level, however, was quite low, ranging from 0.02 to 0.06 % of total rapeseed protein. This level is less than would be expected from the level of 2S albumin mRNA synthesized, suggesting that the low level of BNA may be due to inefficient translation of the chimeric mRNA or incorrect processing and targeting of the protein.

In a similar line of research, De Clercq and co-workers (1990) constructed a chimeric gene using the BNA cDNA under the control of the promoter of the 2S albumin gene 1 (AT2S1) of *Arabidopsis.* This chimeric gene was used to transform tobacco, *Arabidopsis* and canola. Results showed that the protein was stably accumulated in the seeds of all three plants. The expression levels of BNA were very low, only 0.05 to 0.3 % of the total salt-soluble seed protein. The presumable reasons for these low levels of expression were that the promoter was not very strong and that only one or two copies of the chimeric gene were integrated. This was demonstrated by a new transformation experiment with *Arabidopsis* plants using the promoter of the *Arabidopsis* 2S albumin gene 2 (AT2S2), which is highly expressed, and constructing a tandem arrangement of two pAT2S2-BNA-3'nos chimeric genes (Conceição et al., 1994). This resulted in an unexpectedly high increase of the BNA accumulation level up to 1.28 % of total seed protein resulting in a methionine increase up to about 20 % compared to untransformed seeds of *Arabidopsis*. However, transformation of canola with the same chimeric genes resulted in a 60-fold lower level of expression, with no detectable impact on the total methionine content of the seeds (Denis et al., 1995).

The first report on transfer and expression of the 2S albumin from Brazil nut in grain legumes came from Townsend and co-workers (1994). They transformed soybean with the methionine-rich protein under the phaseolin promoter resulting in an accumulation level of the Brazil nut protein up to 8 % of the total transgenic seed protein. However, this contributed to no more than 26 % increase of the methionine content. Apparently, the metabolic demands imposed by expressing high levels of BNA suppressed the formation of some endogenous methionine-containing proteins, e.g. the Bowman–Birk proteinase inhibitor. Further experiments showed that supplying exogenous methionine to cotyledons cultured *in vitro* can reverse these effects (Jung et al., 1997).

The groups of Pickardt (1995) and Saalbach (1995) reported the expression of BNA in narbon bean (*Vicia narbonensis*) under the regulation of the seed-specific *Vicia faba* legume B4 (LeB4)-promoter. In seeds of homozygous lines with single gene inserts, BNA amounted to 1 to 3 % of the total protein and in homozygous lines with a tandem insert approximately 4.8 % of the total protein was the methionine-rich protein. In the latter line, total methionine content was doubled in comparison with the wild-type narbon bean (fig. 1.15) whereas the other amino acids remained nearly unchanged. Although the total methionine content was doubled, the total sulfur content seemed to be similar to that of wild-type seeds. Again, increase of the methionine level occurred at the cost of another endogenous S-compound, the dipeptide γ-glutamyl-S-ethenyl-cysteine (GEC). Subsequent experiments pointed out that GEC acts as a sulfur source in *V. narbonensis* seeds that can be mobilized upon sulfur limitation (Muntz et al., 1997). When sufficient sulfur was supplied to the plants, the level of sulfur in the seeds of transgenic lines increased approximately two-fold (Waddell et al., 1997). Consequently, when transgenic narbon bean lines expressing the 2S albumin from Brazil nut are grown under sulfur limitation, seeds have a decreased GEC content that is inversely proportional to the BNA level. But when there is sufficient external sulfur available, seeds have increased methionine and sulfur levels directly correlated to the amounts of the methionine-rich 2S albumin (Muntz et al., 1998).

The methionine-rich protein from Brazil nut was also introduced in common bean (*Phaseolus vulgaris*) via biolistic methods (Aragão et al., 1999). The transgene was driven by a doubled 35S CaMV promoter and AMV enhancer sequences. The transformed protein could be isolated from transgenic seeds. In two of the five transgenic lines, the methionine content was significantly increased (14 and 23 %) over the values found in untransformed plants. No further data were available.

Some time ago, it was discovered that the potent allergenicity of Brazil nuts for some people is attributable to the BNA protein. Not surprisingly, this allergenicity is maintained in BNA expressed in transgenic seeds (Nordlee et al., 1996). When BNA-containing transgenic seeds are intended for animal consumption, this allergenicity may be of little concern. However, the transgenic protein is likely to be unacceptable in human foods.

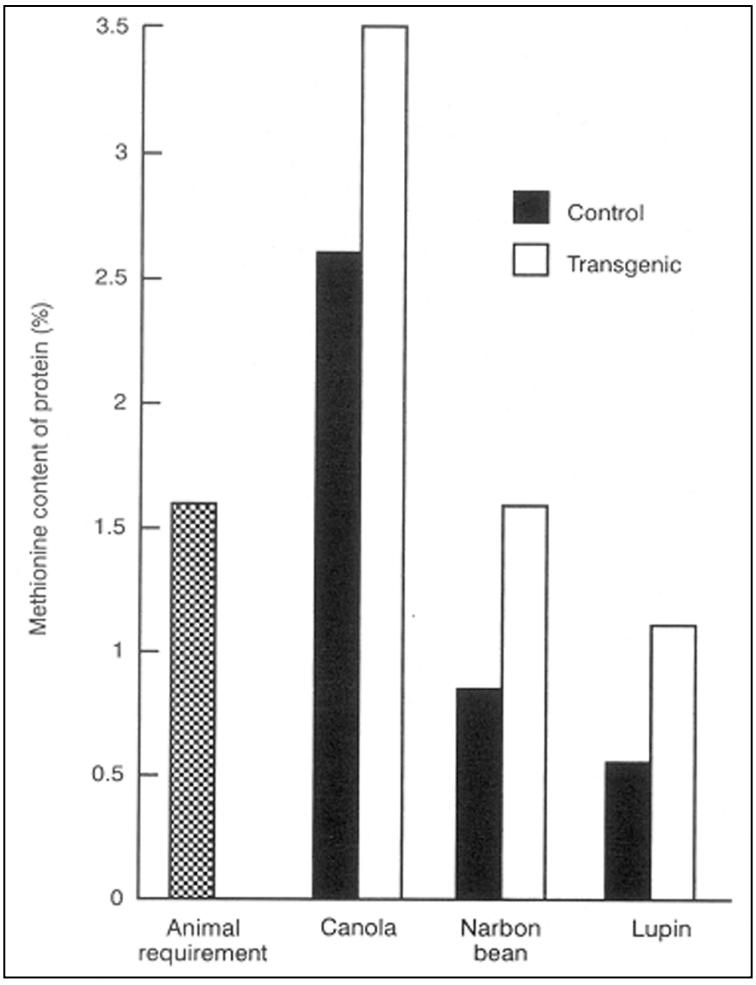


Figure 1.15: Methionine contents of seeds modified using genetic engineering compared with an estimate of the dietary requirement of growing pigs for methionine (taken from Tabe and Higgins, 1998). The pig is chosen as an example of an intensively farmed animal that is a consumer of livestock feeds. The methionine contents are expressed as a percentage of total protein (g of Met per 100 g of protein) for canola (*Brassica napus*, Altenbach et al., 1992), narbon bean (*Vicia narbonensis*, Saalbach et al., 1995), and narrow leaf lupin (*Lupinus angustifolius*, Molvig et al., 1997). The methionine requirement of pigs is expressed as a percentage of dietary protein (g of Met per 100 g of protein). For optimal growth, the pig requires 3.5 % of dietary protein to be composed of S-containing amino acids, of which at least 1.6 % must be methionine. The remaining 1.9 % can be methionine or cysteine (Fuller et al., 1989).

2S albumin of sunflower

Another sulfur-rich protein that has been used for this purpose is the 2S albumin from sunflower, *Helianthus annuus* (SFA) that contains 16 % methionine and 8 % cysteine (Kortt and Caldwell, 1990). A chimeric gene controlled by the promoter of a pea vicilin gene and encoding SFA was introduced into pea and chickpea (Tabe and Higgins, 1998). In transgenic seeds of both plants, SFA accumulated to 2-5 % of the extractable seed protein. This was associated with a significant increase in seed methionine. Surprisingly, the transgenic seeds also had increased amounts of total protein. Therefore, although transgenic pea and chickpea seeds containing SFA had more S-amino acids than controls, their protein was not enriched with respect to these amino acids. The increase in seed protein in the SFA-containing peas and chickpeas was accompanied by a decrease in seed starch content (Tabe and Higgins, 1998).

The same chimeric gene was also introduced into narrow leaf lupin, *Lupinus angustifolius* (Molvig et al., 1997). The methionine-rich protein accumulated to approximately 5 % of total seed protein in a homozygous transgenic line containing a single tandem insertion of the transferred DNA. The only substantial changes in the

amino acid composition of lupin seeds as a result of the expression of SFA were an increase by 94 % of methionine (fig. 1.15) and, unexpectedly, a decrease by 12 % of cysteine. Again, the increase was not in direct proportion to the accumulation level of SFA and this indicated that associated changes must have occurred in the other components of the protein or non-protein fraction of the seed (Tabe and Droux, 2002). Still, the net effect was an increase of 19 % in the S-amino acid content of the seeds. In a feeding trial with rats, it was demonstrated that along with the increased sulfur content of the seeds the biological value of the diet had increased by 15 %. These transgenic seeds were also approved in a feeding trial with chickens. The results documented that the additional methionine was available to the birds and that the transgenic grain had improved nutritive value relative to the parental lupin variety (Tabe and Higgins, 1998).

Because SFA is resistant to degradation in the rumen, it can be used to deliver S-amino acids directly to the small intestine of the animal when introduced into feeding crops. Sheep, in particular, have a high requirement for S-amino acids to fuel the growth of wool, which contains fibre proteins rich in cysteine. Therefore, transgenic lupin seeds containing SFA were used in a feeding experiment with sheep (White et al., 2000). Sheep fed the transgenic lupin grain had an 8 % higher rate of wool growth and 7 % higher live weight gain than sheep fed the parent grain. This is the first report where a transgenic modification increased the nutritive value of a feed grain for a ruminant animal and where it is shown that the effects are consistent with an increase in the flow of methionine to the tissues.

In other experiments, the methionine-rich sunflower protein was introduced in leaves of alfalfa (Tabe et al., 1995), subterranean clover (Khan et al., 1996; Christiansen et al., 2000) and tall fescue (Wang et al., 2001).

More recently, the 2S albumin of sunflower was found to be allergenic too, reducing its usefulness for improving plant nutritional quality (Kelly and Hefle, 2000).

Zeins of maize

Generally, methionine-rich proteins derived from monocotyledonous plants are introduced in transgenic monocot seeds and proteins from dicotyledonous plants are used in transgenic dicot seeds because there is a lot of difference between them. Not only the regulatory elements for gene expression, like promoter sequences and transcription and polyadenylation signals are different (Ueng et al., 1988), but also the codon usage is not identical. Altering these expression elements from monocot to dicot, makes it possible to express zein genes from maize in dicot plants as well.

A first attempt to introduce a maize 15 kDa zein, which contains 11 % methionine and 4.3 % cysteine in a dicotyledonous plant, was performed by Hoffman and co-workers (1987). The zein gene was transferred to tobacco under control of the phaseolin

promoter. They found that transcription and translation occurred in a proper way and that the protein was processed correctly in developing tobacco seeds. Zein accumulation could reach as much as 1.6 % of the total seed protein. The methionine content was not checked.

The same chimeric gene encoding the maize 15 kDa zein protein was expressed in soybean seeds (Dinkins et al., 2001). Two of the four transgenic lines accumulated the zein protein in the seeds resulting in a 12-20 % increase in methionine and 15-35 % increase in cysteine compared to control lines whereas the content of the other amino acids remained unchanged.

The group of Anthony (1997) kept their investigations to monocotyledonous plants. They isolated the gene for the methionine-rich 10 kDa zein from maize and reintroduced it under the control of its own promoter in another maize genotype. Accumulation levels ranged from 0 % to 0.9 % of the total protein content. The highest accumulation level corresponded to an increase of approximately 30 % of the methionine content of the maize kernels.

In the long term, it can be safer to exploit the sulfur-rich prolamins of maize and other cereals, as these crops have been consumed by human and livestock for many centuries without the development of allergies to their prolamin fractions.

Other methionine-rich proteins from plants

Still a lot of other methionine-rich proteins have been identified from diverse plant sources, and some of their genes have been cloned. A few examples are the AmA1 protein from *Amaranthus hypochandriacus* (Raina and Datta, 1992), the 2S proteins from paradise nut (*Lecythis zabucajo*) and cannonball (*Couroupita quianensis*), two members of the Brazil nut family (Zuo and Sun, 1996) and the 10 kDa proteins from seeds of hemp (*Cannabis sativa*, Odani and Odani, 1998).

Methionine-rich proteins from animals

Also animal-derived genes coding for high-methionine proteins can be used to enhance the methionine content of plants. Recently, the bovine β-casein protein was introduced into soybean under the soybean lectin Le1 promoter (Maughan et al., 1999). In the soybean seeds, the size of the recombinant β-casein protein was reduced due to differences in glycosylation between animals and plants but the β-casein protein constituted approximately 0.1 to 0.4 % of the total soluble seed protein. The influence on the methionine content was not examined.

methionine-rich protein	promoter	hostplant	methionine increase in seeds	reference
2S albumin *Bertholletia excelsa*	β-phaseolin	tobacco	10-30 %	Altenbach et al., 1989
	soybean lectin	canola	undetectable	Guerche et al., 1990
	β-phaseolin	canola	up to 33 %	Altenbach et al., 1992
	AT2S1	tobacco, canola, *Arabidopsis*	undetectable	De Clercq et al., 1990
	AT2S2	*Arabidopsis*	up to 20 %	Conceiçao et al., 1994
	AT2S2	canola	undetectable	Denis et al., 1995
	β-phaseolin	soybean	26 %	Townsend et al., 1994
	LeB4	narbon bean	100 %	Saalbach et al., 1995; Pickardt et al., 1995
	doubled 35S	common bean	14-23 %	Aragão et al., 1999
2S albumin *Helianthus annuus*	vicilin	pea, chickpea	significant	Tabe and Higgins, 1998
	vicilin	lupin	94 %	Molvig et al., 1997
10 kDA zein	zein	maize	30 %	Anthony et al., 1997
15 kDa zein *Zea mays*	β-phaseolin	tobacco	not determined	Hoffman et al., 1987
	β-phaseolin	soybean	12-20 %	Dinkins et al., 2001
β-casein	Le1	soybean	not determined	Maughan et al., 1999
synthetic protein	β-phaseolin, β-conglycinin	tobacco	20 %	Keeler et al., 1997

Table 1.2: Overview of the methionine-rich proteins used to increase the methionine content of seeds.

Although introduction of animal-derived high-methionine proteins can help to increase the methionine content of plants, it is doubtful that consumers will accept these proteins in plant products.

Synthetic methionine-rich protein

Instead of looking for methionine-rich proteins in diverse plant sources, it is also possible to create an entirely synthetic gene sequence encoding an artificial protein with a high methionine content. Keeler and co-workers (1997) created a *de novo* protein containing 31 % lysine and 20 % methionine and expressed the gene encoding this protein in tobacco driven by the seed-specific promoter of the bean phaseolin gene or the soybean β-conglycinin gene. Both promoters provided a level of expression in the mature transgenic tobacco seeds that resulted in a significant increase of the total lysine and methionine content, both approximately 20 %. When expressed at a similar level in legume seeds, such a protein could, potentially, have a significant effect on the overall methionine content.

Engineering an endogenous seed protein

Another strategy is to isolate the gene for a naturally occurring, sulfur-poor seed protein and to modify its nucleotide sequence so that it encodes a protein with an increased methionine composition. This can be accomplished by insertion of methionine-rich sequence(s) or by site-directed mutagenesis of existing codons to increase the methionine content (table 1.3). For this approach, a critical task is to select a region of the protein that can be altered without affecting the overall structure, stability, function and other cellular and nutritional properties of the engineered protein: sequence modification that interferes with the signal peptide will inhibit the engineered protein to enter the ER; modifications which involve variation in polar and charged residues might cause solubility changes in the protein that interfere with their association in the ER and assembly in the protein storage vacuoles; altering the glycosylation sites can result in changing the structure and properties of the storage proteins; etc.

A first attempt to engineer a seed storage protein was established by Hoffman and co-workers (1988). The gene coding for phaseolin from *Phaseolus vulgaris* was modified by adding a 45 bp nucleotide sequence encoding a region from the maize 15 kDa zein containing six methionine codons. The peptide was inserted into the third exon of the phaseolin gene, resulting in a modified phaseolin gene with nine methionine codons, the high-methionine (hiMet) phaseolin. This modified phaseolin gene was expressed at the same level as the wild-type gene in seeds of transgenic tobacco as measured by mRNA abundance. However, the hiMet phaseolin accumulated to a much lower

concentration than the wild-type protein, only at 0.2 % of the normal phaseolin. With the unraveling of the 3-dimensional structure of phaseolin (Lawrence et al., 1990), there was found that the 45 bp-insertion was made in a region of importance to the stabilization of the phaseolin trimers. This peptide insertion might have destabilized the trimers, rendering the hiMet phaseolin susceptible to proteolytic degradation in developing seeds (Sun, 1999).

In another study, the region between the 6th and 7th cysteine residues of the 2S albumins from different species was found to vary in both length and sequence (Ampe et al., 1986), suggesting that it may tolerate considerable structural change. This possibility was explored by De Clercq and colleagues (1990), using the 2S albumin gene of *Arabidopsis* (AT2S1) as target. A part of the variable region was deleted and replaced with a sequence containing 11 additional methionine codons. The sequence of the inserted segment was partially derived from the homologous 2S albumin from Brazil nut. This modified 2S gene was transferred into *Arabidopsis*, canola and tobacco. The transgenic seeds of all three species were found to accumulate methionine-enriched 2S albumins at levels ranging from 1 to 2 % of the total seed protein.

Also the 2S protein of paradise nut was engineered (Zuo and Sun, 1996). Within the variable region, three different replacements were generated, which give 62.5, 36.5 and 87.5 % increase in the methionine content of this region. The mutant constructs, together with the wild-type gene were transferred into tobacco under the regulation of the phaseolin promoter. Sun (1999) reported that the mutant proteins were stably accumulated in the seeds but further data were not available.

Phytohemagglutinin (PHA), a lectin found in *Phaseolus vulgaris* seeds, was modified by replacing specific amino acid residues with a methionine residue at positions known to be occupied by methionine residues in homologous lectin proteins (Kjemtrup et al., 1994). Mutant PHA proteins with one, three or four additional methionine residues were found to undergo correct post-translational modifications in transformed cultured tobacco cells. Furthermore, they mentioned that the engineered protein with three mutations accumulated stably in the protein storage vacuoles of the cotyledons of transgenic tobacco seeds. No data were presented concerning the amount of protein accumulation in the seeds or the amount of methionine.

By aligning the amino acid sequences of 12S globulins from various legumes and non-legumes, genetically variable domains of the glycinin protein of soybean could be identified (Wright, 1988). The nucleotide sequences corresponding to each of the variable domains were deleted in different glycinin genes and replaced with a synthetic sequence encoding 4 contiguous methionine residues. The stability of these modified proteins was first tested in *E. coli* (Kim et al., 1990). Some of the modified proteins accumulated as soluble proteins in the cells at high levels. Subsequently, two of the modified genes were expressed in tobacco plants under the 35S CaMV promoter. Both modified glycinins accumulated in the leaves, stems and seeds (Utsumi et al., 1993).

To obtain seed-specific expression, the two modified glycinin genes were expressed in tobacco under the promoter of the rice storage protein glutelin. Consequently, the altered glycinins were targeted to the seed protein vacuoles and accumulated at high levels (1-4 % of total seed proteins). However, about half of the synthesized glycinins was susceptible to limited degradation (Takaiwa et al., 1995). The auteurs mentioned that the degradation was not related to the modification, but to the heterologous nature of the protein.

Modification of the 12S legumin from *Vicia faba* was reported by Saalbach and colleagues (1995). A frame shift mutation was introduced into the legumin B4 (*LeB4*) gene close to the 3' end, resulting in a shorter carboxyl terminus of the protein with two additional methionine residues. The new stop codon was subsequently converted to a lysine codon extending the reading frame by 16 additional nucleotide triplets including two additional methionine codons. All these changes were created in exon 3 of the *LeB4* gene. This exon corresponds to the 3' terminal exon of the homologous gene *Gy2* of soybean coding for glycinin, the 12S storage protein of soybean. In addition, a *Gy2-LeB4* hybrid gene, the glycigumin gene, was constructed by substituting the 3' terminal exon of the soybean *Gy2* gene with the modified 3' terminal exon of the *LeB4* gene. Both modified proteins could be expressed in yeast, but could not be detected in transgenic tobacco seeds when expressed under their own promoter. The conclusion was that these proteins were rapidly degraded. In following experiments, they were able to show that the mutated C-terminus of the proteins interfered with correct folding and opened the polypeptides to premature degradation (Muntz et al., 1997).

In the same publication, the modification of the 7S vicilin from *Vicia faba* was described. Vicilin was engineered by substituting 8 amino acid residues with a methionine residue. Five different positions for replacement were selected from a comparison of homologous 7S globulin sequences from several plant species (Doyle et al., 1996). In addition, three positions were selected where, with a high probability, a conservative amino acid exchange should not affect the functionality of the vicilin. The construct was put under the control of the seed-specific USP-promoter from *Vicia faba* and transformed into tobacco. In this experiment it was possible to detect the altered protein in transgenic tobacco seeds. No significant difference was found between the accumulation levels of the wild-type and mutated vicilin polypeptides in seeds of tobacco transformants (1.25–2 % of the globulin fraction, Saalbach et al., 1995).

In none of the experiments published, the methionine content of the seeds could be increased significantly. This is mainly because the expression levels of the altered genes remained very low. A possible explanation for these low expression levels is that seed proteins are encoded by multigene families and only one of the gene families' members was altered and expressed in the host plant. Another problem was degradation of the modified proteins. The modifications can have changed the structure of the protein rendering it susceptible to degradation in the ER or in the

protein storage vacuoles. Another possibility is that expression in a heterologous host plant harboring another quality control system in the ER and other plant proteases, caused degradation of the seed protein.

protein	modification	promoter	hostplant	result	reference
β-phaseolin *Phaseolus vulgaris*	insertion 6 Met	own	tobacco	unstable	Hoffman et al., 1988
AT2S1 *Arabidopsis thaliana*	insertion 11 Met	own	tobacco, *Arabidopsis*, canola	stable 1-2 %	De Clercq et al., 1990
2S albumin *Lecythis zabucajo*	insertion 3/5/7 Met	own	tobacco	stable	Sun, 1999
PHA *Phaseolus vulgaris*	substitution 1/3/4 Met	own	tobacco	stable	Kjemtrup et al., 1994
glycinin *Glycine max*	insertion 4 Met	glutelin	tobacco	stable 1-4 %	Takaiwa et al., 1995
12S legumin *Vicia faba*	frameshift 4 Met	own	tobacco	unstable	Saalbach et al., 1995
7S vicilin *Vicia faba*	substitution 8 Met	USP	tobacco	stable 1.25-2 %	Saalbach et al., 1995
glycigumin *Glycine max*	hybride 4 Met	own	tobacco	unstable	Muntz et al., 1997

Table 1.3: Overview of the modifications performed with seed storage proteins to increase the methionine content of seeds.

STRATEGY OF THIS WORK

The strategy used in this work to increase the methionine content of *Phaseolus* beans is modification and seed-specific expression of the gene coding for an endogenous seed storage protein, the arcelin-5a protein.

There are several reasons why the arcelin-5a protein was chosen for this purpose. First of all because arcelins are very abundant proteins, both in the wild beans where they contribute 30 to 40 % of the total protein content, as well as in the cultivated beans in which arcelin was introgressed by breeding. Also in seeds from *P. acutifolius* and *A. thaliana* lines transformed with the *arc5-I* gene, high accumulation levels of the arcelin-5a protein were obtained, up to 24 % and 15 % respectively (Goossens et al., 1999). Secondly, because arcelins are only found in very few wild genotypes of *P. vulgaris,* there is no wild-type gene to compete with in the cultivated *Phaseolus* beans, in which the modified *arc5-I* gene will be introduced. Thirdly, there is a low homology between the protein sequences of the different arcelin variants (for example: 63 % homology between arcelin-5 and arcelin-1, Goossens et al., 1994) indicating that modification of the protein sequence is feasible without loss of the stability of the protein. Another good reason is the complete characterization of the gene encoding arcelin-5a as well as the protein itself (Goossens et al., 1994; 1995; 1999a) and the determination of the crystallographic structure of arcelin-5a (Hamelryck et al., 1996a), allowing to predict the influence of mutations by computer modeling. To increase the methionine content of legume seeds substantially by expressing a modified seed storage protein, it is important that the stability of the protein is preserved after modification and that the modified protein can be accumulated at a sufficient level in the seeds. The listed qualities of arcelin-5a made the protein a very good candidate for the chosen strategy.

To modify the arcelin-5a protein, substitutions and insertions with methionine residues were made with site-directed mutagenesis. Four different constructs with 3 or 4 methionine substitutions at predetermined positions were designed. In addition, two constructs with a methionine-rich insertion in a variable loop were created. The stability of these modified proteins was tested experimentally by transformation of *A. thaliana*.

To increase the methionine content of the arcelin-5a protein as much as possible, all mutations resulting in a stable methionine-enhanced protein were combined in several possible combinations. Another five different constructs were designed and the stability of these modified proteins was tested in *A. thaliana* seeds in turn.

The final goal of this work is to contribute to the development of a strategy to increase the methionine content of *Phaseolus* beans. Although stable transformation of most grain legumes remains difficult to achieve, an *Agrobacterium*-mediated transformation procedure for *P. acutifolius* was established (Dillen et al., 1997). In order to use this transformation method to transfer the modified arcelin-5a proteins into *Phaseolus* beans, the protocol was optimized. The effect of different factors on T-DNA transfer was examined by measuring transient expression levels of an intron-containing *uidA* gene. Finally, this improved transformation procedure could be used to introduce three modified arcelin-5a proteins with 4, 7 and 12 additional methionines into *Phaseolus* beans.

MODIFICATION OF ARCELIN-5A

the results of this chapter will be submitted together with the results of chapter 3 as 'De Clercq J., Hamelryck T., Goossens A., Wyns L., Loris R., Van Montagu M., Angenon G. (2002). Enhancement of the methionine content of the *Phaseolus vulgaris* seed protein arcelin-5 by protein engineering.'

Introduction

Several biotechnological strategies to increase the methionine content of seeds have been described in literature. One strategy is engineering the seed amino acid metabolism in order to increase the free amount of methionine. This can, for example, be obtained by expressing a mutant feedback-insensitive gene coding for a key enzyme of the methionine biosynthetic pathway. Seed-specific expression of a bacterial desensitized aspartate kinase, the first enzyme of the aspartate pathway, increased the production of free methionine two- to three-fold in seeds of tobacco (Karchi et al., 1993) and *Vicia narbonensis* (Pickardt et al., 1998). A second approach is seed-specific expression of heterologous genes coding for high-methionine proteins. The gene coding for the 2S albumin from Brazil nut (*Bertholletia excelsa),* which contains 18.8 % methionine, has been used widely for this purpose. Seed-specific expression of this gene resulted in significant increases of the methionine content in seeds of *Nicotiana tabacum* (Altenbach et al., 1989), *Arabidopsis thaliana* (Conceiçao et al., 1994) *Brassica napus* (Altenbach et al., 1992), *Vicia narbonensis* (Pickardt et al., 1995), *Glycine max* (Townsend et al., 1994) and *Phaseolus vulgaris* (Aragão et al., 1999). Another sulphur-rich protein is the 2S albumin from sunflower (*Helianthus annuus,* 16 % methionine) that was successfully introduced in *Lupinus angustifolius* (Molvig et al., 1997). However, these two proteins revealed to be allergenic and their use can result in the development of allergenic reactions in humans or livestock (Nordlee et al., 1996; Kelly and Hefle, 2000). Most probably, their use will be abandoned in the future. Other methionine-rich proteins used to increase the methionine content of seeds are the zeins from *Zea mays*. Accumulation of the 15 kDa zein in seeds of *Glycine max* resulted in significant enhancement of the methionine content (Dinkins et al., 2001). The 10 kDa zein was successfully introduced in maize kernels to improve the methionine concentration (Anthony et al., 1997). Beside the naturally occurring high-methionine proteins, a *de novo* protein containing 30 % methionine was designed and successfully produced in seeds of *Nicotiana tabacum* (Keeler et al., 1997). A third possible strategy to increase the methionine content of seeds is engineering and seed-specific expression of a gene coding for an endogenous seed storage protein: isolating the gene for a methionine-poor seed protein and

protein/gene	host plant	promoter	level[a]	reference
2S1 albumin *Arabidopsis thaliana*	*Nicotiana tabacum*	own	0.1	De Clercq et al. (1990)
2S albumin *Bertholletia excelsa*	*A. thaliana*	at2S1	1.0-2.0	De Clercq et al. (1990)
	Brassica napus	at2S1	1.0-2.0	De Clercq et al. (1990)
	"	soybean lectin	0.02-0.06	Guerche et al. (1990)
	"	β-phaseolin	1.7-4.0	Altenbach et al. (1992)
	N. tabacum	β-phaseolin	3.0-8.0	Altenbach et al. (1989)
	"	at2S1	1.0-2.0	De Clercq et al. (1990)
	Vicia narbonensis	CaMV 35S	0.0-0.01	Saalbach et al. (1994)
	"	legumin B4	1.0-4.8	Pickardt et al. (1995)
2S albumin *Helianthus annuus*	*Lupinus angustifolius* *Pisum sativum*	pea vicilin own	5.0 2.0-5.0	Molvig et al. (1997) Tabe and Higgins (1998)
β-conglycinin *Glycine max*	*Petunia hybrida*	own	0.1-1.0	Beachy et al. (1985)
glycinin *G. max*	*N. tabacum* *Oryza sativa*	glutelin GluB1 glutelin GluB1	1.0-4.0 0.1-5.0	Takaiwa et al. (1995) Katsube et al. (1999)
lectin *G. max*	*N. tabacum*	own	0.2	Okamuro et al. (1986)
β-phaseolin *P. vulgaris*	*N. tabacum* *O. sativa*	own glutelin Gt1	1.0 0.0-4.0	Sengupta-Gopalan et al. (1985) Zheng et al. (1995)
phytohemagglutinin-L *P. vulgaris*	*N. tabacum*	own	0.02-0.05	Sturm et al. (1988)
α-Amylase inhibitor *P. vulgaris*	*P. sativum*	PHA-L	1.0-3.0	Schroeder et al. (1995)
Psl lectin *P. sativum*	*N. tabacum*	own	0.2-0.9	de Pater et al. (1996)
legumin A *P. sativum*	*N. plumbaginifolia* *O. sativa* *Triticum aestivum*	own glutelin Gt1 LTW glutelin	0.2 0.0-4.2 0.3-1.5	Ellis et al. (1988) Sindhu et al. (1997) Stöger et al. (2001)
vicilin *Vicia faba*	*N. tabacum*	USP	2.2	Saalbach et al. (1995)
vicilin *P. sativum*	*N. tabacum*	own	0.5	Higgins et al. (1988)
10 kDa Zein *Zea mays*	*Zea mays*	own	0.0-0.9	Anthony et al. (1997)
15 kDa Zein *Z. mays*	*N. tabacum* "	β-phaseolin CaMV 35S	0.02-1.6 1.0-2.0	Hoffman et al. (1987) Bagga et al. (1995)

Table 2.1: Examples of heterologous expression of genes encoding seed storage proteins (modified from Goossens et al., 1999a).

[a] Protein level, as percentage of total extractable protein in transgenic seeds.

modifying its nucleotide sequence so that it encodes a protein with an enhanced methionine composition. So far, the methionine content could not be increased significantly by using this strategy due to low expression levels of the modified gene or degradation of the modified protein (Hoffman et al., 1988; De Clercq et al., 1990; Kjemtrup et al., 1994; Saalbach et al., 1995; Takaiwa et al., 1995; Muntz et al., 1997; Sun, 1999). A fourth strategy is influencing the accumulation level of the seed proteins by using the anti-sense approach. By repressing the accumulation of methionine-poor seed proteins, other methionine-richer proteins can be produced in higher amounts. In *Brassica napus,* the seed protein cruciferin was repressed in favor of napin what resulted in a small increase of the methionine content (Kohno-Murase et al., 1995).

Our goal was to find a successful strategy to improve the methionine content of legume seeds in general and *Phaseolus* beans in particular. In most of the cases, the level at which an introduced heterologous seed protein accumulates is rather low, usually not more than a few percent (table 2.1). Therefore, we decided to modify a leguminous protein, the *Phaseolus vulgaris* seed storage protein arcelin-5a. Arcelins are abundant seed storage proteins found in some wild *P. vulgaris* genotypes. So far, seven arcelin variants have been identified, of which the arcelin-5 variant is used here. Arcelin-5 consists of two major polypeptides: arcelin-5a (Arc5a) and arcelin-5b (Arc5b) encoded by the *arc5-I* and *arc5-II* genes respectively (Goossens et al., 1994). The Arc5a protein consists of 240 amino acids, none of which is a methionine. This protein was chosen because (1) the protein is accumulated at very high levels in seeds; (2) the gene encoding it as well as the protein itself are well characterized and (3) the crystallographic structure has been determined (Hamelryck et al., 1996a), allowing to predict the influence of mutations by computer modeling.

The methionine content of Arc5a was increased by substituting certain candidate amino acids with a methionine codon and by inserting a methionine-rich loop at the surface of the protein. First, six modified *arc5-I* genes were created with 3 or 4 methionine substitutions or one insertion of a methionine-rich loop. In a second step, mutations resulting in a high-accumulating protein were combined to obtain the highest possible methionine content of Arc5a.

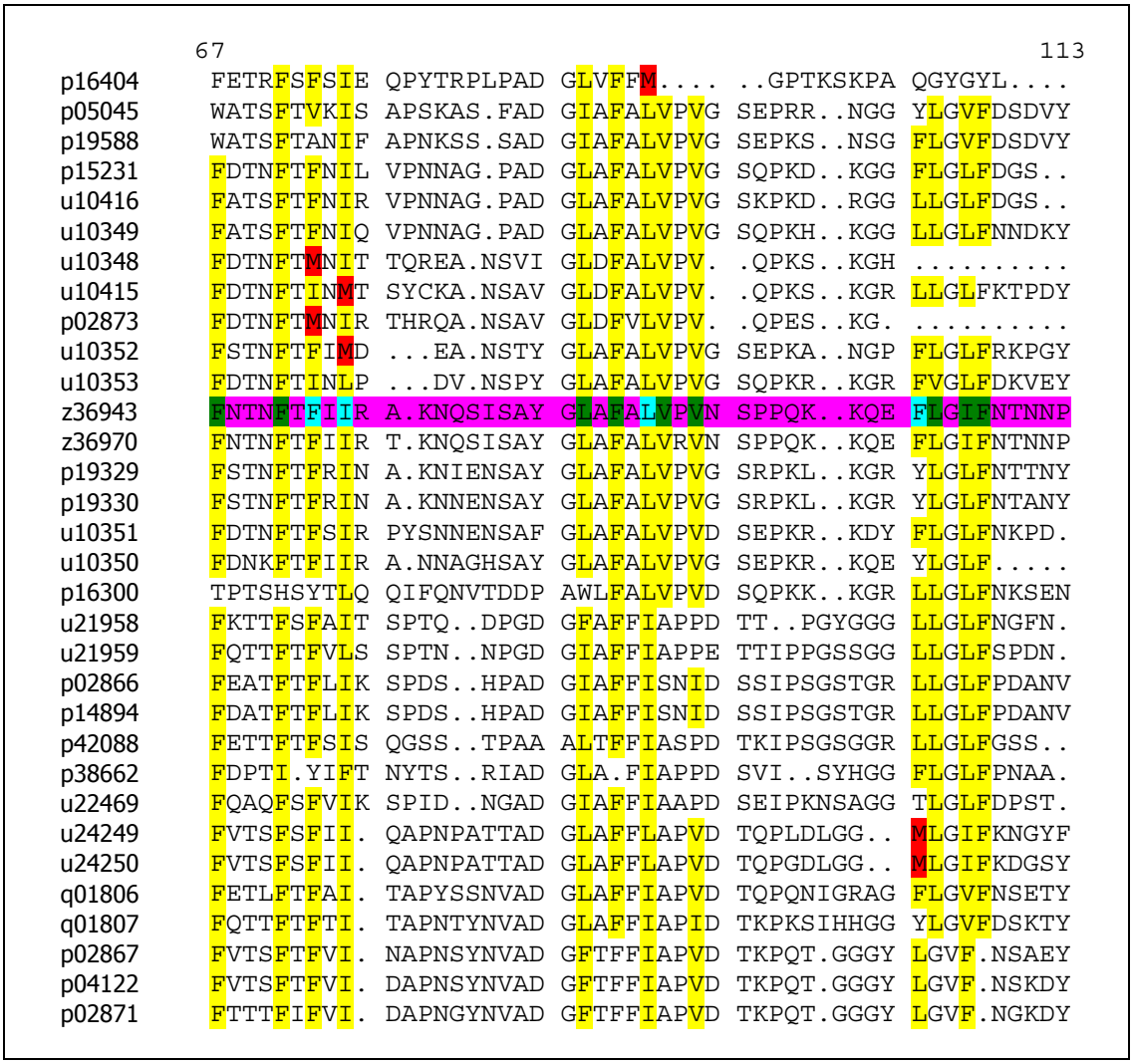


Figure 2.1: Part of the alignment of the amino acid sequences of Arc5a (pink) with homologous leguminous proteins, from amino acid 67 to amino acid 113. Retained candidate amino acids for substitution are indicated in green when mostly hydrophobic residues (Leu, Ile, Val and Phe, indicated in yellow) could be found at this position in the homologous proteins and in blue when one or more methionines (indicated in red) and hydrophobic residues could be found at the indicated position. Only the most related proteins are shown: p16404 (lectin precursor of *Erythrina corallodendron*), p05045 (phytohemagglutinin-L of *Phaseolus vulgaris*), p19588 (lectin DB58 precursor of *Dolichos biflorus*), p15231 (phytohemagglutinin-M precursor of *P. vulgaris*), u10416 (phytohemagglutinin-L of *P. acutifolius*), u10349 (phytohemagglutinin-E of *P. vulgaris*), u10348 (α-amylase inhibitor-2 of *P. vulgaris*), u10415 (α-amylase inhibitor-like lectin of *P. acutifolius*), p02873 (α-amylase inhibitor-1 precursor *of P. vulgaris*), u10352 (α-amylase inhibitor-3 of *P. vulgaris*), u10353 (α-amylase inhibitor-like lectin of *P. maculatus*), z36943 (arcelin-5a of *P. vulgaris*), z36970 (arcelin-5b of *P. vulgaris*), p19329 (arcelin-1 of *P. vulgaris*), p19330 (arcelin-2 of *P. vulgaris*), u10351 (arcelin-4 of *P. vulgaris*), u10350 (arcelin-like protein of *P. acutifolius*), p16300 (lectin precursor of *P. lunatus*), u21958 (lectin-1 of *Cladrastis kentukea*), u21959 (lectin-2 of *C. kentukea*), p02866 (Concanavalin A precursor of *Canavalia ensiformis*), p14894 (Concanavalin A precursor of *Canavalia gladiata*), p42088 (lectin of *Bowringia milbibraedii*), p38662 (lectin of *Dolichos lablab*), u22469 (lectin of *Arachis hypogaea*), u24249 (lectin-1 precursor of *Robinia pseudoacacia*), u24250 (lectin-2 precursor of *R. pseudoacacia*), q01806 (lectin-1 precursor of *Medicago truncatula*), q01807 (lectin-2 precursor *of M. truncatula*), p02867 (lectin precursor of *Pisum sativum*), p04122 (lectin-chains of *Lathyrus ochrus*), p02871 (lectin of *Vicia faba*).

Results

Modification of the Arc5a protein through substitutions

The *arc5-I* gene was modified by substituting several amino acid codons for a methionine codon. The sites for these replacements were selected from a comparison of amino acid sequences of more than seventy proteins of the legume family that are homologous to the Arc5a protein, including the different arcelin variants, α-amylase inhibitors, phytohemagglutinins and other lectins and lectin-like proteins (fig. 2.1).

The strategy for selection of candidate amino acids to be substituted was as follows:

First, all positions in the Arc5a amino acid sequence where a leucine, isoleucine, valine or phenylalanine was found, were surveyed. These amino acids are the best candidates for replacement with a methionine because they are also hydrophobic and most similar to methionine. Also in modification experiments with β-phaseolin, Dyer and colleagues (1993, 1995) experienced that substitution of these amino acids with a methionine resulted in the smallest possible destabilization of the protein. These positions were examined in the alignment taking into account which amino acids were localized at the corresponding positions in the homologous proteins. When there were one or more methionines and/or mostly hydrophobic amino acids situated at the investigated position in the homologous proteins, the amino acid of Arc5a at the examined position was retained as a candidate for substitution. For example, an isoleucine is found in the Arc5a amino acid sequence at position 75; two of the homologous proteins contain a methionine at the corresponding position and most other homologous proteins a hydrophobic residue (see fig. 2.1). Ile75 was thus taken as a candidate amino acid for substitution.

Using this strategy, 52 candidate amino acids could be identified.

In a next step, the location of these candidate amino acids in the three-dimensional structure of the Arc5a protein was explored (table 2.2). This was possible because the crystallographic structure of Arc5a is available (Hamelryck et al., 1996a). Also the relative accessibility of the total side chain of the candidate amino acid was calculated (table 2.2). This relative accessibility describes the surface of the amino acid that is exposed to the solvent (water in our case).

In the hydrophobic core of the protein, the packing of amino acids is very dense. An amino acid situated in this core is hardly accessible and subject to many interactions with other hydrophobic residues. Replacing one of these amino acids can result in destabilization of the protein. In contrast, an amino acid situated at the surface of the protein is easily approachable and has no or much less interaction with neighboring amino acids. Most probably, these amino acids can be replaced without any problem. Therefore, most candidate amino acids situated in the hydrophobic core were rejected.

Two exceptions were made. The first one was for the amino acids Phe73 and Ile75. At these positions, the residues were successfully substituted with a methionine in phytohemagglutinin (Kjemtrup et al., 1994).

amino acid	accessibility (%)	location
Leu17	30.3	S+++
Leu41	52.1	S+++
Val44	38.3	S+++
Leu47	10.4	S+
Ile58	1.6	HC+
Phe73	0.2	HC+++
Ile75	0.1	HC+++
Leu91	2.4	HC+++
Val94	41.9	S+++
Phe104	11.2	S+++
Ile138	14.2	S+
Phe148	0.2	HC+++
Leu190	2.2	HC+
Val194	0.1	HC+
Leu219	26.3	S+++

Table 2.2: Relative accessibility in % of the total side chain of the candidate amino acids and location of these amino acids in the crystallographic structure of Arc5a: at the surface (S+++), at the surface but less approachable (S+), in the hydrophobic core (HC+++) or near the hydrophobic core (HC+). Only candidate amino acids that were used to create the constructs are mentioned.

Another exception was made for the amino acids Leu91 and Phe148. During computer analysis, superposition of several crystallographic structures of leguminous lectins (see insertions) on the structure of Arc5a was performed. In this superposition, all positions where a methionine was found in one of the leguminous lectins were investigated. If the structure of the lectin around the methionine residue was very similar or equal to the structure of Arc5a, the amino acid located at this position in Arc5a was taken as a candidate amino acid. For example, in the protein sequence of the lectin of *Erythrina corallodendron* (Ecorl) a methionine is found at position 95. When the 3D structure of this lectin is superposed on the 3D structure of Arc5a, this position corresponds to Leu91 in Arc5a. As can be seen in figure 2.2, the structure of the two proteins is very similar at this position. So, Leu91 is sustained as a candidate for substitution although it is located in the hydrophobic core. The position of Phe148 in Arc5a corresponds to Met42 in concanavalin A of *Canavalia ensiformis* and both proteins have a similar structure at this position (not shown).

After rejecting all candidate amino acids with a low accessibility located in the hydrophobic core, 26 candidates remained.

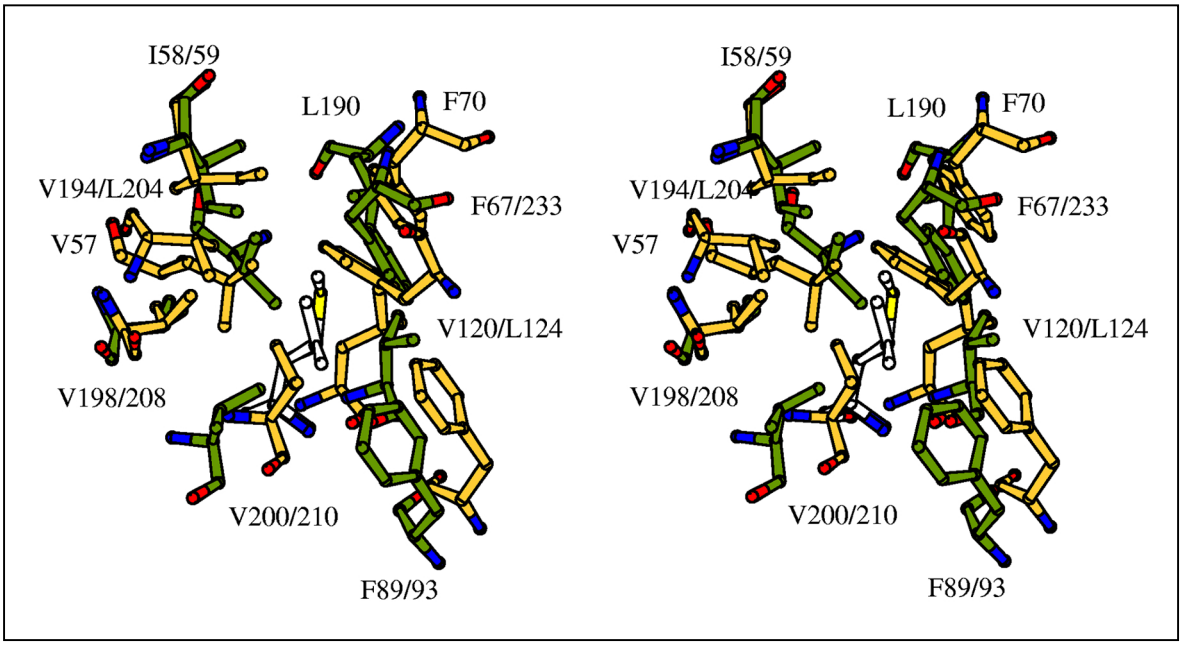


Figure 2.2: Superposition of the region around Leu91 in Arc5a on the corresponding region in Ecorl around Met95. The superposition was done using the C_α, C_β and C_γ atoms of the Leu91 and Met95 residues (both are shown in white). Arc5a residues are shown in green and labeled first, Ecorl residues are shown in yellow. All residues with side chain atoms within a radius of 6 Å of the C_γ atom of Met95 are visualized.

Thirdly, the feasibility of a methionine replacement at the different positions in Arc5a was evaluated via computer modeling. The candidate residues in the Arc5a structure were mutated *in silico* to methionine, and subjected to energy minimization (steepest descent minimization, followed by conjugate gradient refinement). In most cases, the energy of the mutant after refinement was close to the wild type energy (table 2.3).

amino acid	delta E	structure maintenance
Leu17	- 0.9	+++
Leu41	- 2.1	+
Val44	- 6.0	+++
Leu47	- 4.1	+++
Ile58	- 3.7	+
Phe73	- 6.9	+
Ile75	- 6.5	+++
Leu91	- 9.0	+++
Val94	- 2.3	+++
Phe104	- 7.2	+++
Ile138	- 6.3	+++
Phe148	- 5.7	+++
Leu190	+ 2.2	+
Val194	- 13.4	+++
Leu219	- 2.1	+++

Table 2.3: Difference in minimal E between the mutant and wild-type structure (delta E, kcal/mol) and maintenance of the structure after methionine replacement: no changes (+++) or very little change (+). Only candidate amino acids that were used to create the constructs are mentioned.

After refinement, the structure of the mutant was compared to the wild-type structure. Preservation of the structure (main chain conformation and side chain conformation) was then taken as the final criterion to accept or reject a potential replacement position (table 2.3).

In this way, five candidates were rejected because the structure after substitution deviated from the original structure.

Of the 21 amino acid candidates that fulfilled all three requirements, the phenylalanines were omitted because their structure differs the most with the structure of methionine. Again, a few exceptions were made. From structure analysis it was known that a small hydrophobic cluster is formed near the surface by Phe104 and Phe108, and that Ile138 is situated close to this cluster (fig. 2.3). *In silico* mutation to methionine and energy minimization of Phe104 and Ile138 suggested that the simultaneous replacement of these residues with methionine would preserve the hydrophobic character of the site while preserving a reasonable geometry. Therefore, the simultaneous replacement of Phe104 and Ile138 with methionine was suggested as a reasonable choice. Also Phe73 and Phe148 were sustained for reasons described above.

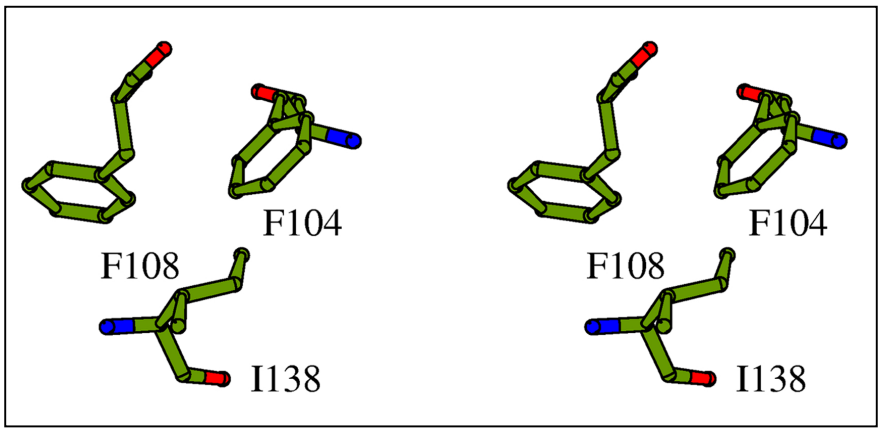


Figure 2.3: Stereo figure of the Phe104 / Phe108 / Ile138 cluster. Phe104, Phe108 and Ile138 are shown as ball-and-stick models. Carbon atoms are shown in green, nitrogen atoms in blue and oxygen atoms in red.

Finally, four modified Arc5a proteins were designed with each 3 to 4 substitutions:

► in the modified **Arc5a-M1** protein the amino acids Leu17, Leu41, Val44 and Leu47 are substituted with a methionine;

► in the **Arc5a-M2** protein Ile58, Phe73 and Ile75 are substituted;

► in the **Arc5a-M3** protein Leu91, Val94, Phe104 and Ile138 are replaced;

► in the **Arc5a-M4** protein Phe148, Leu190, Val194 and Leu219 are substituted.

The positions of the modifications in the three dimensional structure of Arc5a are shown in figure 2.4.

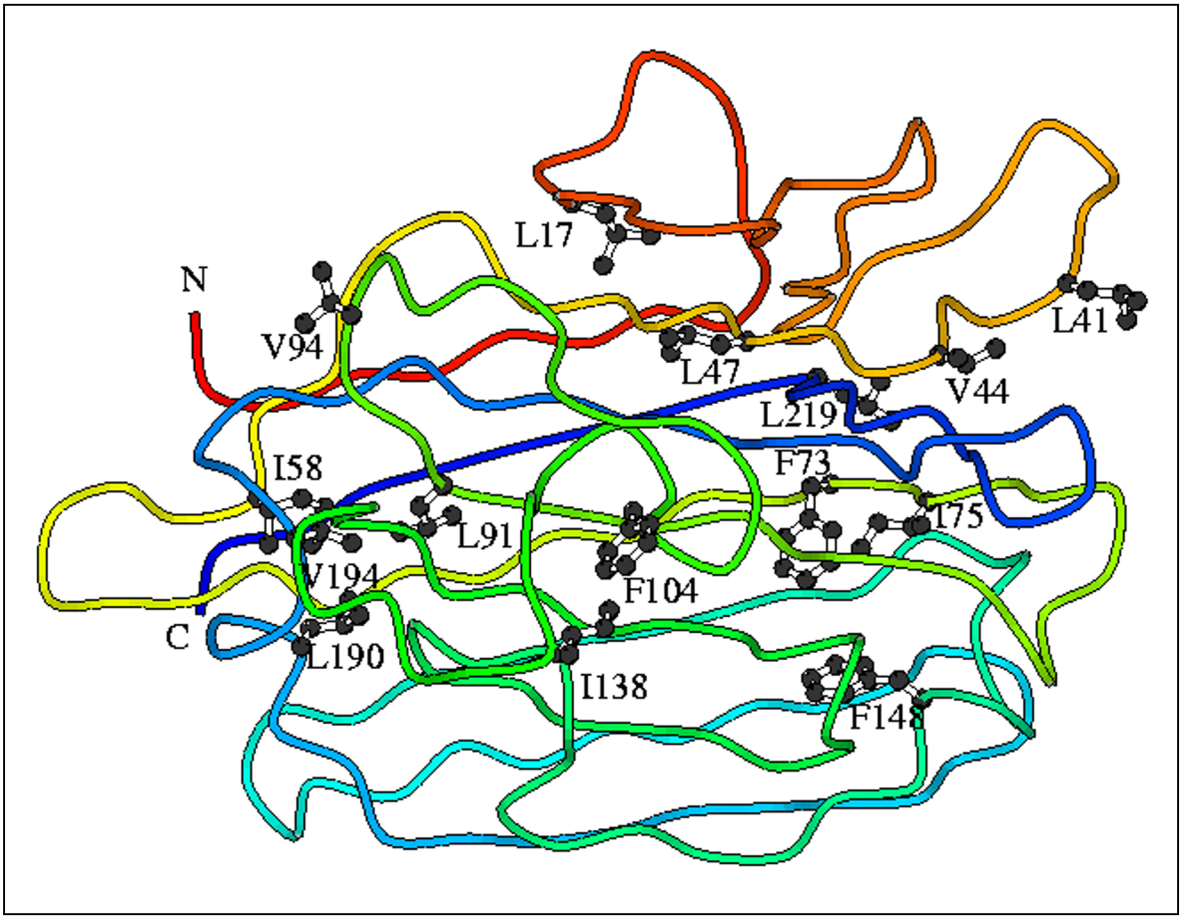
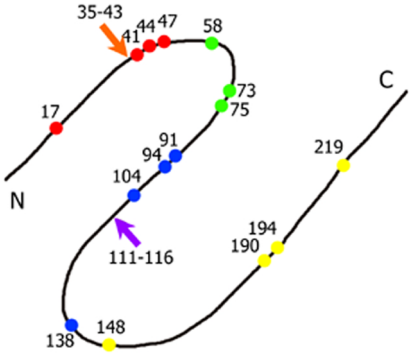


Figure 2.4: Positioning of all substitutions used to create the constructs in the Arc5a structure. The distribution of the substitutions over the four different constructs is indicated in the scheme on the right. The amino acids that are mutated in the modified Arc5a-M1 protein are indicated in red, Arc5a-M2 in green, Arc5a-M3 in blue and Arc5a-M4 in yellow. The positioning of the two insertions is also indicated in the scheme, Arc5a-MI1 in orange and Arc5a-MI2 in purple. C = carboxyl terminus of the protein, N = amino terminus of the protein.



All substitutions were performed using site-directed mutagenesis on the coding sequence of *arc5-I* with the help of several oligonucleotides (table 2.4A). When designing the mutagenic oligonucleotides for methionine replacement, the length of the primer, the stability of the ends of the primer, the composition of the nucleotides, and several other parameters were taken into account (Piechocki and Hines, 1994). Moreover, additional mutations were created when a nucleotide sequence showed similarities with a potential polyadenylation site (AATAAA or AATAAT) (Dean et al., 1986; Joshi, 1987) or with the sequence ATTTA, which is responsible for destabilization of the mRNA in plants (Shaw and Kamen, 1986). Sequences that could be recognized as a plant intron, were avoided as well (van Aarssen et al., 1995). If the

mutations mentioned above did not create or destroy a restriction site, an additional mutation was incorporated to allow detection of the mutated genes by restriction analysis. When designing these additional mutations, the codon usage in *Phaseolus* plants was taken into account. To calculate this codon usage, nucleotide sequences of *Phaseolus* genes with a high expression level (arcelin-1, arcelin-4, arcelin-5 and phaseolin) were used.

The final constructs used for plant transformation were created in a step-by-step procedure (see experimental procedures). All steps were controlled by restriction analysis. The introduced substitutions were verified by sequencing after mutagenesis and after the final cloning step.

Modification of the Arc5a protein through insertion

Instead of substituting amino acid codons for a methionine codon, inserting a methionine-rich sequence is another approach to increase the methionine content of the arcelin-5a protein. This can be done at the C-terminus of the protein, or in a variable loop at the surface of the protein. Young and co-workers (1999) published that Arc5a undergoes post-translational processing in which 8 to 11 amino acids are clipped from the C-terminus. This event was expected because no electron density is observed for the terminal 12 amino acids when studying the 3D structure of Arc5a (Hamelryck et al., 1996a). Therefore, we chose to insert a methionine-rich sequence in a variable loop. To identify these variable loops, a superposition of the crystallographic structures of Arc5a and 10 other legume lectins was made: pea lectin (Einspahr et al., 1986); lentil lectin (Loris et al., 1993); *Lathyrus ochrus* isolectins I (Bourne et al., 1990) and II (Bourne et al., 1994); *Griffonia simplicifolia* lectin IV (Delbaere et al., 1989); *Erythrina corallodendron* lectin (Shaanan et al., 1991); concanavalin A (Becker et al., 1975); peanut agglutinin (Banerjee et al., 1994); soybean agglutinin (Dessen et al., 1995) and phytohemagglutinin-L (Hamelryck et al., 1996b). This superposition revealed several loops at the surface of the proteins that varied in size between the different proteins. In this way, five different variable loops of Arc5a where an insertion should be possible could be identified (fig. 2.5):

35 Gly-Val-Gly-Ser-Asn-Glu-Leu-Pro-Arg 43
59 Lys-Asp-Ser-Asn-Asn-Val 64
78 Lys-Asn-Gln-Ser-Ile 82
111 Asn-Asn-Pro-Glu-Pro-Asn 116
207 Gly-Leu-Thr-Glu-Asp-Thr 212

Out of these five possibilities, the loops Gly35-Arg43 and Asn111-Ans116 were chosen because these are the largest and most variable loops of the protein.

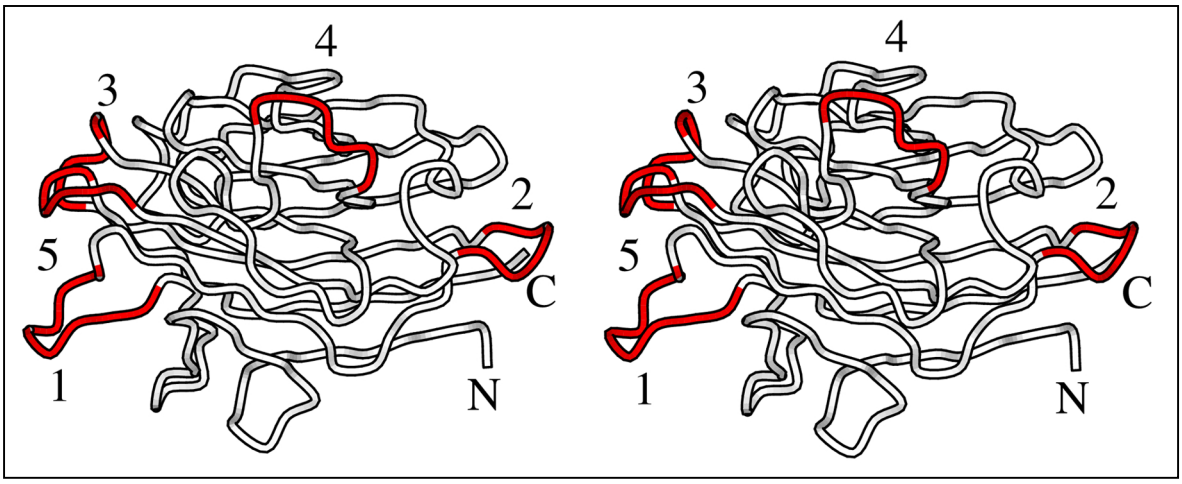


Figure 2.5: Variable loops at the surface of the Arc5a protein identified through superposition of the crystallographic structure of Arc5a with 10 other legume lectins. Loop 1: Gly35-Arg43, loop 2: Lys59- Val64, loop 3: Lys78-Ile82, loop 4: Ans111-Ans116, loop 5: Gly207-Thr212.

There are several possible approaches to modify a loop. The strategy chosen here was to delete the original loop and to replace it with a methionine-rich loop.

When designing a high-methionine loop, several criteria have to be considered. The sequence must be methionine-rich, hydrophilic, favor the formation of a looping sequence, and prevent the nucleation of alternative secondary structures. Dyer and co-workers (1993) composed a 15 amino acid sequence that fulfilled all these criteria:

Asp^--Met-Lys^+-Gly-Met-Met-Asn-Lys^+-Asp^--Met-Pro-Met-Asn-Asp^--Ser.

To maintain hydrophilicity, a total of five charged residues were incorporated to compensate the hydrophobicity of the five methionine residues in the insert. Both acidic and basic residues were selected to preserve iso-electric charge and to promote loop formation by electrostatic interactions (Lys3 and Asp14). Glycine and proline residues were incorporated into the insert because glycine provides enhanced flexibility and proline is favorable for turn conformation.

This 15 amino acid sequence was used as methionine-rich loop to insert in the Arc5a protein.

The two Arc5a proteins with a methionine-rich insertion are:

- ► the **Arc5a-MI1** protein where the loop Gly35-Pro42 was substituted;
- ► the **Arc5a-MI2** protein with replacement of the loop Asn111-Asn116.

The methionine-rich sequences were inserted in the coding sequence of *arc5-I* using site-directed mutagenesis with the help of two long oligonucleotides (table 2.4B). When designing these primers, the same criteria were used as for the primers for single codon substitutions. After mutagenesis and after the final cloning step, the presence of the insertion was verified by sequencing.

Modification of the Arc5a protein by combining groups of mutations

After testing the stability of the modified proteins in *Arabidopsis thaliana* seeds (see chapter 3), the groups of mutations generating highly accumulating methionine-enhanced proteins were combined in five additional modified Arc5a proteins. This was done to increase the methionine content of the Arc5a protein as much as possible.

The retained groups of mutations were the substitutions of Arc5a-M1 (4 Met) and Arc5a-M2 (3 Met), both resulting in a stable methionine-rich Arc5a protein that was produced at levels even higher than unmodified Arc5a *in A. thaliana* seeds, and the insertions of Arc5a-MI1 (5 Met) and arc5a-MI2 (5 Met), both resulting in high accumulation levels of partially degraded proteins (see chapter 3). Although the modified Arc5a proteins with an insertion were not accumulated as a full-length protein in *A. thaliana* seeds, it is likely that all protein products are still present in the seeds and can thus contribute to an increase of the methionine content of the seeds.

The modified Arc5a-M4 protein accumulated only at low levels and Arc5a-M3 could not be detected in the *A. thaliana* seeds.

The proteins combining groups of mutations that were created are the following:

► the **Arc5a-M12** protein with the substitutions of Arc5a-M1 and Arc5a-M2;

► the **Arc5a-M12I1** protein, combining the substitutions of Arc5a-M1 and Arc5a-M2 together with the insertion of Arc5a-MI1;

► the **Arc5a-M12I2** protein, combining the substitutions of Arc5a-M1 and Arc5a-M2 together with the insertion of Arc5a-MI2;

► the **Arc5a-M2I12** protein, combining the substitutions of Arc5a-M2 together with the insertions of Arc5a-MI1 and Arc5a-MI2;

► the **Arc5a-M12I12** protein, combining all retained mutations.

To create these modified genes different strategies were used. In some cases, new site-directed mutagenesis reactions with several (new) oligonucleotides (table 2.4-2.5) were conducted. Other modified genes were created using different cloning strategies (see experimental procedures). Again, all steps were controlled by restriction analysis and in the final step, the combined mutations were verified by sequencing.

Discussion

During past attempts, several different strategies have been used to increase the methionine content of seeds. One useful strategy is to transfer a methionine-rich seed storage protein gene from one species to another. A possible problem with expressing a seed storage protein in a heterologous system is that the transferred protein accumulates at low levels in the heterologous plant even with the use of a strong seed-specific promoter (table 2.1). This can be partially explained by the fact that seed storage proteins are encoded by multigene families, and an individual gene therefore only contributes to a fraction of the total seed storage protein. Another reason can be the lack of essential *trans*-acting factors for seed-specific expression present in the appropriate amounts or at the appropriate time in the heterologous plant. Another problem that may occur is proteolytic degradation of the seed storage protein in the heterologous system. For example, phaseolin from *P. vulgaris* is not as stable in tobacco seeds as it is in bean seeds. Bustos and colleagues (1991) found that only one third of the polypeptides were present as full-length molecules in tobacco seeds.

We decided to engineer a leguminous seed protein by substituting several amino acids with a methionine or by inserting a methionine-rich sequence. When applying this strategy, the two same important conditions mentioned above must be fulfilled: (1) The stability of the protein must be preserved in order to protect the protein from degradation in the ER or in the protein storage vacuoles (PSV). Therefore, the protein's ability to be properly transported, processed, and deposited in the storage cells of seeds must remain unchanged. (2) The modified protein must accumulate at high levels in the seeds.

Although the arcelin-5 protein is encoded by only two genes per haploid genome (Goossens et al., 1994), the accumulation level in wild genotypes of *P. vulgaris* reaches 30 to 40 % of the total seed protein. When the unmodified Arc5a protein was transferred to *P. acutifolius*, accumulation levels in transgenic seeds reached up to 25 % of the total extractable protein content (Goossens et al., 1999a). These very high accumulation levels made the *P. vulgaris* seed storage protein arcelin-5a a suitable candidate for modification and expression in *Phaseolus* beans. Moreover, in a heterologous system, namely in transgenic *A. thaliana* seeds, accumulation levels up to 15 % of the total extractable protein content were established (Goossens et al., 1999a). This suggests that the arcelin-5a protein can be used in a broad range of plants, including non-legumes.

To preserve the stability of the modified protein, a good location of the mutations is essential. Hoffman and co-workers (1988) dramatically destabilized the protein β-phaseolin of *P. vulgaris* when inserting a methionine-rich loop without knowledge of the three-dimensional structure of the protein. Analyzing the 3D structure of β-phaseolin revealed that the insertion was made in the middle of an α-helical domain

(Lawrence et al., 1990). This region is of importance to the stabilization of the phaseolin trimers and the insertion might have destabilized the trimers, rendering the modified protein susceptible to proteolytic degradation (Sun, 1999). In contrast, after analyzing the tertiary structure, the groups of De Clercq (1990) and Tu (1998) successfully modified the 2S albumin of *A. thaliana* and *Bertholetia excelsa*, respectively. They made modifications in a variable loop of the protein between conserved structural domains. Based on comparisons with other members of the 2S albumin family, they concluded that this region could tolerate a variety of amino acids.

Before starting the modification of the Arc5a protein, it was possible to examine the sites for mutation very carefully. The sites for substitution were chosen on the basis of an alignment of 78 leguminous proteins homologous with Arc5a. This method proved to be successful before when it was used to locate mutation sites in phytohemagglutinin of *P. vulgaris* (Kjemtrup et al., 1994), the glycinin protein of *Glycine max* (Utsumi et al., 1993) and the 7S vicilin from *Vicia faba* (Saalbach et al., 1995). In a following step, the location of the candidate amino acids for substitution in the Arc5a three-dimensional structure was investigated, eliminating most sites located in the hydrophobic core of the protein. Finally, methionine replacements were simulated using computer modeling.

Although computer simulations checked every mutation, some combinations of substitutions still destabilized the Arc5a protein. Most probably, destabilization caused by a particular group of substitutions is related to only one or two of these mutations. The Arc5a-M4 protein (substitutions at Phe148, Leu190, Val194 and Leu219) accumulated only at very low levels in *A. thaliana* seeds (see chapter 3). A possible explanation is the location of the amino acids Leu190 and Val194. Both are situated close to each other near the hydrophobic core. Substituting the two together may have rearranged the order within the protein. Another possibility is that substitution of Phe148 within the hydrophobic core caused destabilization of the protein, making the modified Arc5a-M4 susceptible for degradation. Substitutions in the hydrophobic core were the most risky ones and the structure of phenylalanine diverges most of the structure of methionine.

The modified Arc5a-M3 protein (substitutions at Leu91, Val94, Phe104 and Ile138) could not be isolated from the transgenic *A. thaliana* seeds although Southern analysis showed that the transgene was present in the transformants (see chapter 3). As mentioned before, we assumed that substituting Phe104 and Ile138 together was favorable to preserve the hydrophobic character and the geometry of the site. But it's possible that substituting the two together has created a solvent exposed hydrophobic patch that resulted in expression problems due to misfolding, decreased stability or aggregation. Again an amino acid in the hydrophobic core (Leu91) was substituted, which might cause problems. Furthermore, there are almost no data available yet about the sequences needed for transport, folding and deposition of the legume seed storage proteins. Consequently, it is also possible that the modifications interfered

with one of these sequences, resulting in a protein that could not pass the quality control system of the ER or that was susceptible to degradation in the protein storage vacuoles.

The sites for insertion were selected by searching variable loops at the surface of the proteins using the superposition of several 3D structures of leguminous lectins. We assume that this strategy to look for possible insertion places was reliable. The influence of the insertion on the complete structure of the protein, on the other hand, was unpredictable. Both modified Arc5a proteins harboring a methionine-rich insertion underwent one or more cleavages or other modifications in *A. thaliana* seeds. Therefore, we assume that the loop itself caused problems. The influence of inserting the designed loop in a seed protein was never tested before in plants. It is possible that the folding of the designed loop changed the distance between the two anchor-points of the loop too much, resulting in disturbance of the total protein structure and thus making the Arc5a protein more susceptible to degradation. It's also possible that the loop contains a cleavage site for plant proteases. Till now, the nature of the degradation products could not yet be determined. An alternative strategy that might circumvent some of the encountered problems would be to use a natural occurring methionine-rich loop for substituting the variable loops of the arcelin-5a protein.

Of the modified proteins combining groups of mutations, all proteins harboring the substitutions of group M2 and/or M1 together with the insertion situated in the loop Gly35-Arg43 (Arc5a-M12I1, Arc5a-M2I12 and Arc5a-M12I12), could not be detected in transgenic *A. thaliana* seeds (see chapter 3). Most probably, the combination of these mutations created problems for the stability of these modified proteins. A possible explanation is that the hydrophobic side-chains of the methionines interacted because they were situated very close to each other, resulting in a changed conformation of the proteins and thus making them unstable.

The strategy we followed was based on a careful choice of substitution sites based on several criteria. Nevertheless, these remain predictions that have to be tested experimentally. Testing all mutations individually is very time consuming and combining many substitutions might be risky. The chosen strategy of making several constructs with 3 to 4 substitutions appears to achieve a good balance between those extremes.

Experimental procedures

Tracing sites for mutation

To select the substitution sites of the arcelin-5a protein, the amino acid sequences of Arc5a with 78 homologous proteins of the legume family were aligned using the PILEUP program of the GCG package (Genetics Computer Group, Madison, WI). The homologous proteins were found in different databases looking for 'lectin' and 'Fabaceae'. Not only amino acid sequences were taken into account but also nucleotide sequences were used after translation with the program TRANSLATION of the GCG package.

Examination of the structure of Arc5a and of the location of the different amino acids in the structure was done with the program O (Jones et al., 1991). This program was also used to create the methionine mutants *in silico*. In all cases, the methionine rotamer closest to the rotamer of the wild-type amino acid was chosen. The relative accessibility of the total side chain of the candidate amino acids was calculated with the program NACCESS (Hubbard and Thornton, 1993). Molecular modeling of the methionine mutants was performed on a Silicon Graphics Personal IRIS workstation with the Insight II/Discover programs (Biosym, Inc., USA). Energy minimization of the mutants was done using simulated annealing refinement with the ESFF force field. Hydrogen atoms and atomic charges were generated with the BIOPOLYMER module. All C_α atom positions were tethered to their initial positions in order to prevent large shifts in the backbone positions. Initial bad contacts and strains were removed using conjugate gradient refinement. Simulated annealing was done by heating the protein to 1000 K over 2000 fs, and keeping the protein at that temperature for another 2000 fs. The protein was cooled to 300 K in 1000 s, and finally minimized with conjugate gradient refinement. The same procedure was also applied to the wild-type structure, in order to calculate the energy difference between the energy-minimized wild-type and *in silico* mutant structures.

To determine which regions of the Arc5a protein would best tolerate an insertion, the crystallographic structures of 10 different legume lectins were superposed on the three-dimensional structure of arcelin-5a, using the CCP4 program LSQKAB (Collaborative Computational Project, 1994).

Figures were designed using the program MOLSCRIPT (Kraulis, 1991).

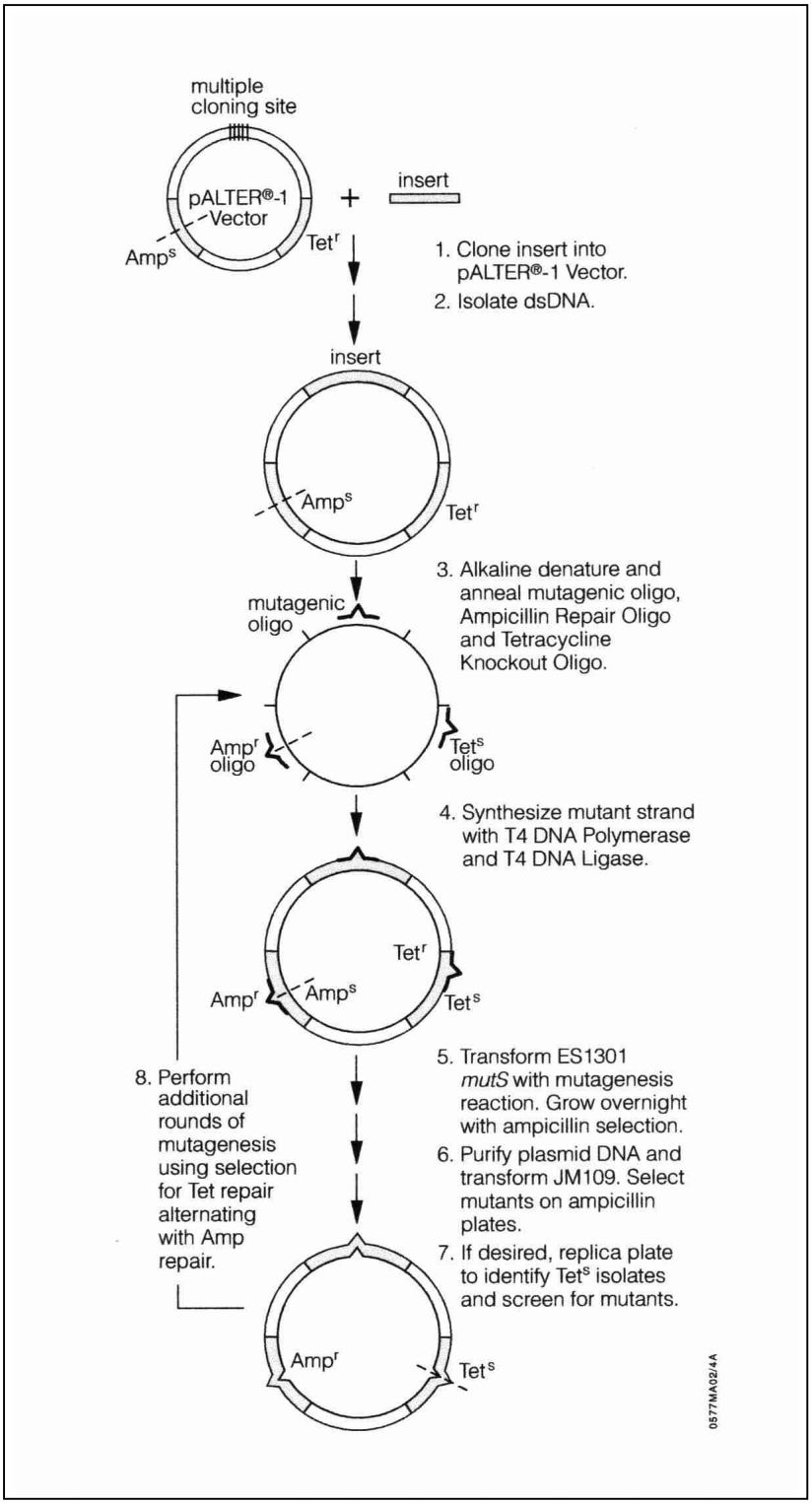


Figure 2.6: Schematic diagram of the 'Altered Sites® II *in vitro* Mutagenesis Systems' procedure.

Site-directed mutagenesis

Site-directed mutagenesis was performed using the 'Altered Sites® II *in vitro* Mutagenesis Systems' procedure (fig. 2.6; Promega, Madison, WI). All steps were carried out according to the guidelines of the manufacturer. Briefly, a *SacI/XbaI* fragment containing the *arc5-I* coding sequence (fig. 2.7) was cloned into the pALTER®-1 vector (fig. 2.8) provided by the manufacturer for production of single-stranded DNA and site-directed mutagenesis. For the production of single-stranded template, an *E. coli* JM109 culture containing the pALTER®-1 vector with the *arc5-I* gene was infected with R408 helper phage. Subsequently, the mutagenesis reaction was started, involving annealing of the antibiotic repair oligonucleotide (provided by the manufacturer) and the mutagenic oligonucleotides to the DNA template, followed by synthesis of the mutant strand with T4 DNA polymerase and T4 DNA ligase. The heteroduplex DNA was then transformed into the repair minus *E. coli* strain ES1301 *mutS* to avoid selection against the desired mutations. A subsequent transfer into the *E. coli* strain JM109 ensured proper segregation of the mutant and wild-type plasmids and resulted in a high proportion of mutants. All mutagenic oligonucleotide sequences used to create the different constructs are listed in table 2.4 and 2.5. To design the mutagenic primers, the OLIGO program was used to increase the stability of the complex between primer and template. All mutagenic primers were synthesized on an oligonucleotide synthesizer (Applied Biosystems, Foster City, CA).

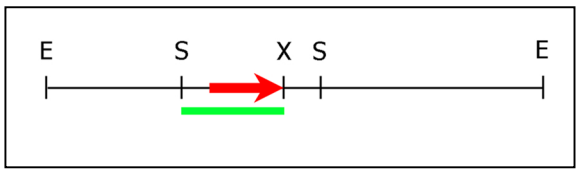


Figure 2.7: Scheme of the genomic *EcoRI* fragment containing the coding sequence of *arc5-I* (red arrow) with indication of the *SacI/XbaI* fragment used for mutagenesis (green) and the restriction sites *EcoRI* (E), *SacI* (S) and *XbaI* (X).

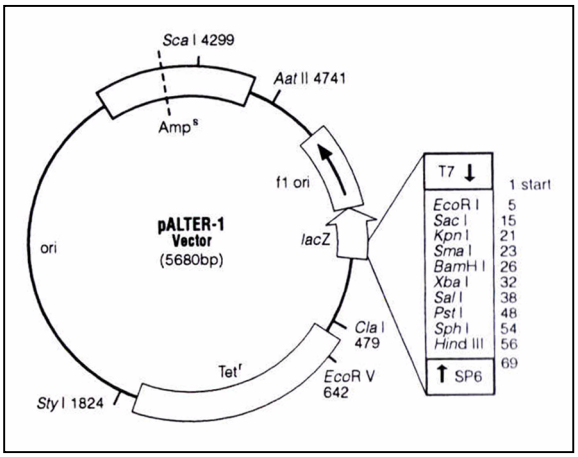


Figure 2.8: pALTER®-1 vector with the ampicilline (Amp) and tetracycline (Tet) resistance genes, of which the Amp gene is mutated (-----), an origin of replication (ori) and a phage origin of replication (f1 ori), and the lacZ gene with a multicloning site sequence in the promoter, flanked with the T7 and SP6 RNA polymerase transcription initiation sites.

Preparation of the constructs harboring *arc5-I* with a few mutations

To create the first six modified *arc5-I* genes, the *SacI/XbaI* fragment containing the *arc5-I* coding sequence derived from the plasmid pATARC3-B1b (Goossens et al., 1999a) was inserted in the pALTER®-1 vector for site-directed mutagenesis. All mutagenic oligonucleotide sequences used to create the six different constructs are listed in table 2.4.

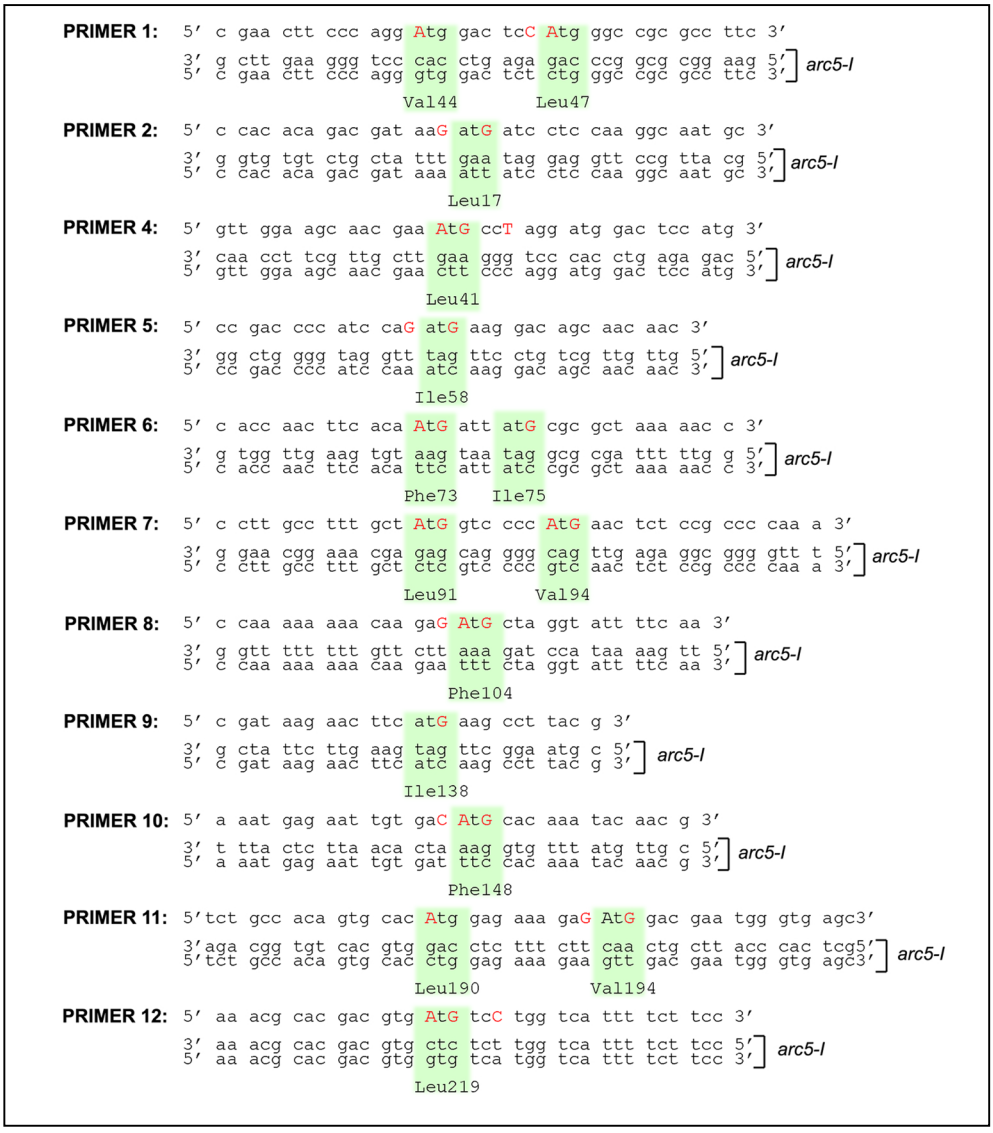


Table 2.4A: Different mutagenic oligonucleotides used to create substitutions with methionine codons. Primer 1, 2 and 4 were used in two subsequent mutagenesis reactions to create the mutations of *arc5I-M1*; primer 5 and 6 were used in a single reaction to create *arc5I-M2*; primer 7, 8 and 9 were used together in one reaction to create *arc5I-M3* and primer 10, 11 and 12 were used in a single reaction to create *arc5I-M4*. Red capital letters represent replaced nucleotides, green boxes contain the triplets of the amino acids substituted with a methionine.

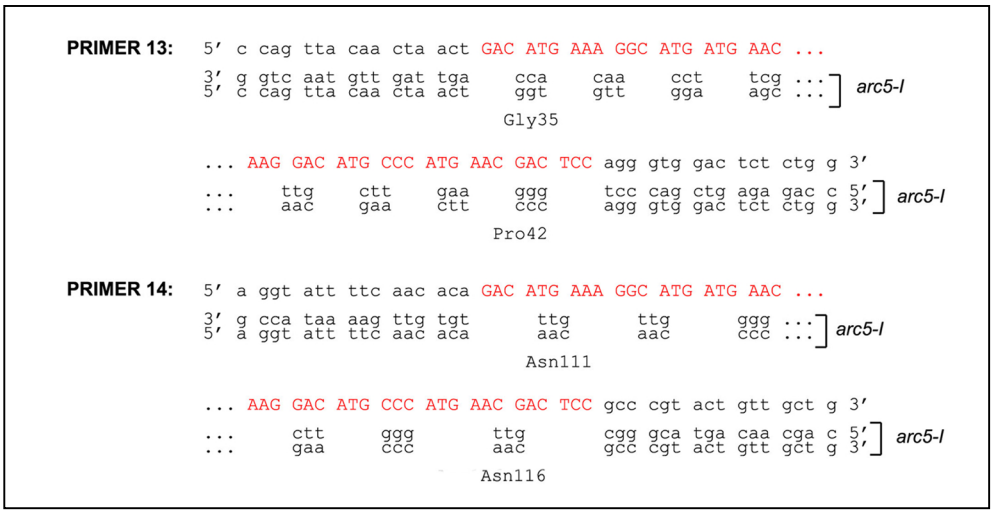


Table 2.4B: Different mutagenic oligonucleotides used to create insertions with a methionine-rich loop. Primer 13 was used to delete the loop Gly35-Pro42 and substitute it with a methionine-rich sequence to create *arc5I-MI1* and using primer 14 the loop Asn111-Asn116 was substituted to create *arc5I-MI2*. Red capital letters represent the methionine-rich sequence.

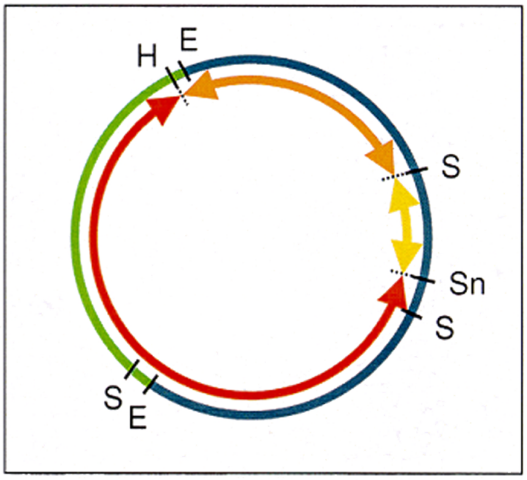


Figure 2.9: Scheme of the complete genomic *arc5-I* clone (blue) in the pBluescriptKS vector (green) with indication of the three fragments used in the ligation: *SacI/SnaBI* (yellow) with the desired mutations, *SnaBI/HindIII* (red) and *HindIII/SacI* (orange).

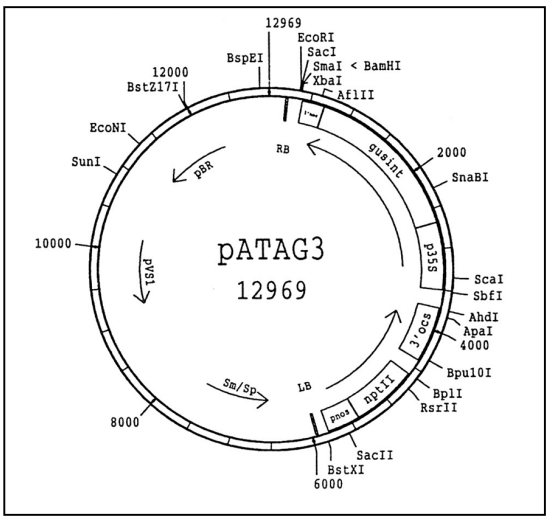


Figure 2.10: Plasmid map of the binary vector pATAG3 containing between the T-DNA borders the *nptII* gene under control of the nopaline synthase (*pnos*) promoter and the octopine synthase 3′ termination and polyadenylation signals *(3′ocs*) and the *Escherichia coli gus* gene (Jefferson, 1987) with the potato *st-ls1* intron (*gusint*; Vancanneyt et al., 1990) under control of the CaMV 35S promoter (*p35S*) and the nopaline synthase 3′ processing and polyadenylation signals (*3′nos*). pBR, origin of replication; pVS1, stability and replication functions of the *Pseudomonas aeruginosa* pVS1 plasmid (Deblaere et al., 1987); Sm/Sp, spectinomycin and streptomycin resistance locus; RB and LB, right and left border repeat of the T-DNA. Single-cutting restriction enzymes are shown outside of the plasmid.

After mutagenesis, the *arc5-I* coding sequence with mutations was cloned in the *EcoRI* fragment of *arc5-I* (fig. 2.7) located in the pBluescriptKS vector (Strategene, La Jolla, CA, USA) using a three-fragment-ligation (fig. 2.9). This *EcoRI* fragment represents the largest *arc5-I* sequence available from the genomic clone (Goossens et al., 1995), harboring potential TATA and CCAT boxes, the majority of the potential *cis*-regulatory elements for seed-specific expression and three potential 5' MAR sequences (see chapter 1). Finally, this *EcoRI* fragment was cloned into the *EcoRI* restriction site of the binary vector pATAG3 (fig. 2.10, Goossens et al., 1999a).

Preparation of the constructs harboring *arc5-I* with several groups of mutations

To combine the retained groups of mutations into five new constructs, different strategies were used. For the *arc5I-M12* gene, the pALTER®-1 vector harboring the *SacI/XbaI* fragment of *arc5I-M1* was taken as starting material for mutagenesis with primers 5 and 6 (table 2.4A). The subsequent steps were as before. The *arc5I-M12I1* gene was produced starting from the pALTER®-1 vector harboring the *SacI/XbaI* fragment of *arc5I-M12*, performing mutagenesis with a new primer 15 (table 2.5). The subsequent steps were as before. To create *arc5I-M12MI2*, other fragments were used for a three-fragment-ligation: (1) the *HindIII/SacI* fragment from the pBluescriptKS vector containing the genomic fragment of *arc5-I*, (2) the *Sau96/HindIII* fragment from the pALTER®-1 vector containing the *SacI/XbaI* fragment of *arc5I-MI1* and (3) the *SacI/Sau96* fragment from pALTER®-1 harboring the *SacI/XbaI* fragment of *arc5I-M12*. For the *arc5I-M2I12* gene, the pALTER®-1 vector containing the *SacI/XbaI* fragment of *arc5I-M2* was taken as starting material for mutagenesis with primers 13 and 14 (table 2.4B). The subsequent steps were as before. To create the *arc5I-M12I1*2 gene, the pALTER®-1 vector harboring the *SacI/XbaI* fragment of *arc5I-M12* was used for mutagenesis with primer 14 and 15 (table 2.4B-2.5).

Table 2.5: Mutagenic oligonucleotide used to create a methionine-rich insertion in the loop Gly35-Arg43, taking the substitutions of the amino acids Val44 and Leu47 with a methionine codon into account. Red capital letters represent the methionine-rich sequence.

Bacterial growth

The *E. coli* strains ES1301*mutS*, JM109 and MC1061 were used for plasmid manipulations. Bacteria of MC1061 were grown at 37 °C on solidified (1.5 % agar) or in liquid LB medium (10 g/l bacto tryptone, 0.5 g/l yeast extract, 5 g/l NaCl), or in liquid SOC medium (2 g/l bacto tryptone, 0.5 g/l yeast extract, 10 mM NaCl, 2.5 mM KCl, 20mM Mg^{2+}, 20 mM glucose) after transformation, bacteria of ES1301*mutS* in liquid LB medium and bacteria of JM109 on solidified or in liquid LB medium, in liquid SOC medium after transformation or on minimal M9 medium (6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g/l NaCl, 1g/l NH_4Cl, 15 g/l agarose, 2mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2 % glucose, 1 mM thiamine-HCl) before single-stranded DNA preparation, always supplemented with appropriate antibiotics. For transformation, heat-shock competent cells of ES1301*mutS* and JM109 (Promega) and electro-competent cells of MC1061 (self-prepared) were used.

DNA sequence analysis

Initial DNA sequences were obtained using the universal SP6 and T7 primers. To complete the sequences on both strands, gene-specific primers were synthesized on an oligonucleotide synthesizer (Applied Biosystems). DNA sequence data were assembled and analyzed using the GCG package (Genetics Computer Group). The percentage of identity and similarity between sequences was determined with the GAP program.

STABILITY OF THE MODIFIED ARCELIN-5A PROTEINS IN ARABIDOPSIS THALIANA SEEDS

the results of this chapter will be submitted together with the results of chapter 2 as 'De Clercq J., Hamelryck T., Goossens A., Wyns L., Loris R., Van Montagu M., Angenon G. (2002). Enhancement of the methionine content of the *Phaseolus vulgaris* seed protein arcelin-5 by protein engineering.'

Introduction

Figure 3.1: An *Arabidopsis thaliana* plant.

Arabidopsis thaliana is a small dicotyledonous plant belonging to the *Brassicaceae* or Mustard family (fig. 3.1) that is well suited for biological research because the plants mature rather quickly, going from germinating seeds, formation of a rosette plant, bolting of the main stem, flowering and production of seeds in siliques in approximately 8-10 weeks. *Arabidopsis* grows very well in the laboratory or the greenhouse where it takes up little space because of its small size. It undergoes self-fertilization, which gives homozygous plants thus allowing genetic traits to be more

easily propagated. *Arabidopsis* is easy to transform through *Agrobacterium tumefaciens*-mediated gene transfer and easy to cross, allowing to introduce and to breed in new characteristics and to study the inheritance.

Furthermore, it was demonstrated before that the unmodified *arc5-I* gene can be expressed in *A. thaliana* seeds at high levels, up to 15 % of the total extractable protein (table 3.1). SDS-PAGE and Immunoblot analysis of the crude protein extracts showed that this unmodified Arc5a protein isolated from transgenic *A. thaliana* seeds co-migrated with the Arc5a protein that was purified from *Phaseolus vulgaris* beans (fig. 3.2). These experiments proved that transcription, translation, posttranslational modification and vacuolar targeting occur in a very similar way in both species.

Line	Copies	Arc5 Protein Level	
		Hemizygous	Homozygous
E-307	1	7.9 ± 0.8	8.6 ± 2.0
E-313	1	7.3 ± 0.6	9.8 ± 2.3
E-105	1	9.5 ± 0.9	11.9 ± 0.1
E-103	1	8.1 ± 0.3	14.7 ± 0.1
E-305	2	9.6 ± 0.4	14.1 ± 2.7
E-102	≥3	4.9 ± 0.3	10.8 ± 4.2
E-306	≥3	7.3 ± 2.4	13.0 ± 1.2

Table 3.1: Characterization of transgenic plants harboring one transgenic locus (derived from Goossens et al., 1999a). Transgenic lines contain the complete genomic clone of the unmodified *arc5-I* gene. The Arc5a protein level is indicated as percentage of total extractable protein of seeds of hemizygous or homozygous transgenic plants. Values are followed by the SD. Measurements were performed on an aliquot of approximately 500 seeds. The seeds harvested from a hemizygous plant were a mixture of homozygous transformed, homozygous non-transformed and hemizygous seeds.

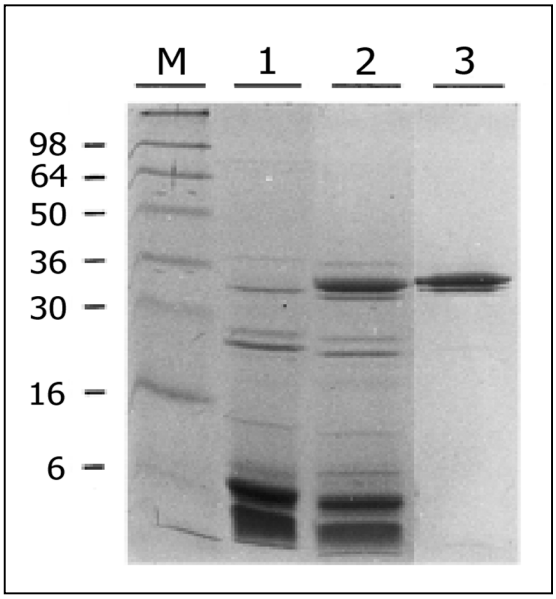


Figure 3.2: SDS-PAGE on crude protein extracts of *A. thaliana* seeds of a non-transformed Colombia-O genotype (lane 1), and of the transformed line E-103 (lane 2). Lane 3 contains arcelin-5 proteins purified from the *P. vulgaris* genotype G02771. Lane M contains the marker proteins (molecular masses are indicated on left in kDa) (derived from Goossens et al., 1999b).

All these characteristics made *Arabidopsis thaliana* an ideal model plant to test the stability of the modified arcelin-5a proteins in the seeds.

Results

Transformation of the modified *arc5-I* genes into *Arabidopsis thaliana*

The modified *arc5-I* genes were transformed into the *A. thaliana* genotype Colombia-O via *Agrobacterium tumefaciens*-mediated gene transfer. Therefore, the different modified genes (fig. 3.3) were inserted between the left and right T-DNA border of the binary vector pATAG3 (Goossens et al., 1999a) under control of their own seed-specific promoter and transferred to the *A. tumefaciens* strain C58C1Rif^R(pPM90). At least 25 plants were transformed for each of the first six modified *arc5-I* genes and 15 plants for each of the combined modified *arc5-I* genes. All transgenic plants had a normal phenotype and produced seeds in similar quantities as wild-type plants.

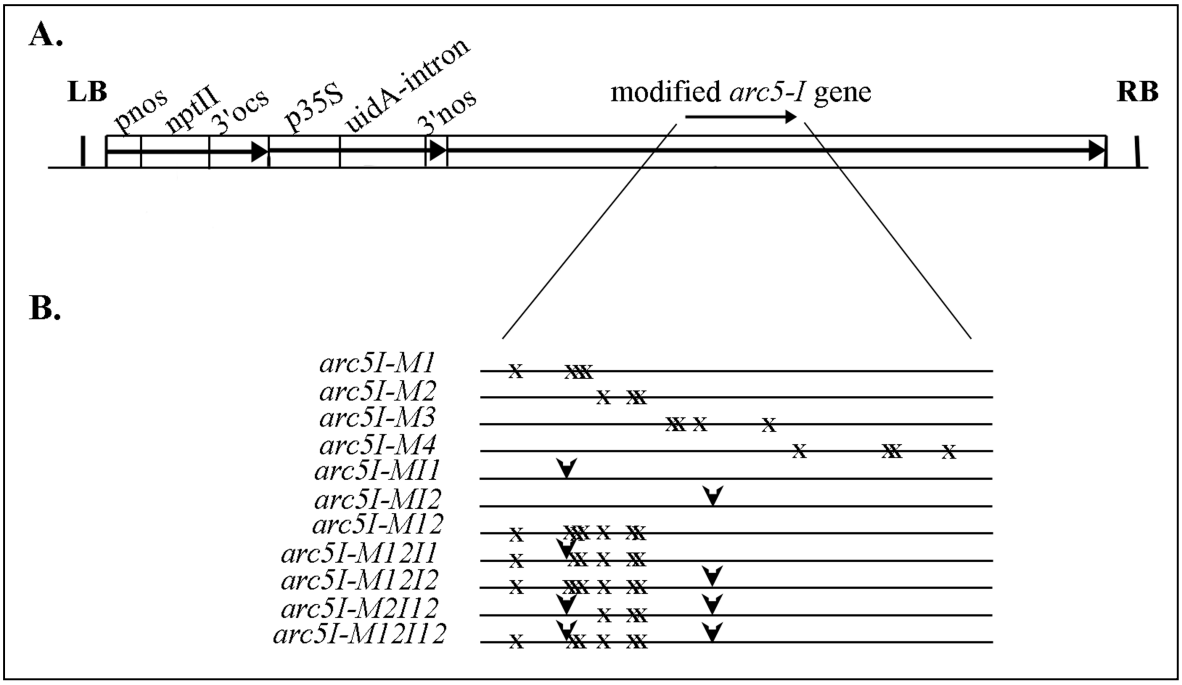


Figure 3.3: (A) The transferred region of the binary vectors with between the left (LB) and right (RB) T-DNA borders the two marker genes neomycin phosphotransferase II (*nptII*) under control of the nopaline synthase promoter (*pnos*) and the octopine synthase 3' termination and polyadenylation signals (*3'ocs*) and the *Escherichia coli* ß-glucuronidase gene with the potato *st-ls1* intron (*uidA-intron*) under control of the CaMV 35S promoter (*p35S*) and the nopaline synthase 3' processing and polyadenylation signals (*3'nos*). In addition, they contain a *P. vulgaris* genomic fragment that codes for a modified Arc5a protein with enhanced methionine content. (B) Indication of the location of the substitutions with a methionine codon (x) and insertions with a methionine-rich sequence (arrow) in the different modified *arc5-I* genes. In *arc5I-M1* the codons for Leu17, Leu41, Val44 and Leu47 are substituted; in *arc5I-M2* the codons for Ile58, Phe73 and Ile75; in *arc5I-M3* the codons for Leu91, Val 94, Phe104 and Ile138; *in arc5I-M4* the codons for Phe148, Leu190, Val194 and Leu219; in arc5I-MI1 the loop Gly35-Pro42 is substituted; in *arc5I-MI2* the loop Ans111-Ans116 is substituted; in *arc5I-M12* the codons of M1 and M2 are substituted; in *arc5I-M12I1* the codons of M1 and M2 and the loop of I1 are substituted; in *arc5I-M12I2* the codons of M1 and M2 and the loop of I2 are substituted; in *arc5I-M2I12* the codons of M2 and the loops of I1 and I2 are substituted; in *arc5I-M12I12* the codons of M1 and M2 and the loops of I1 and I2 are substituted.

Detection of the modified Arc5a proteins in transgenic seeds

Experiments in which the unmodified *arc5-I* gene was expressed in *A. thaliana* plants, showed a very low plant-to-plant variability (Goossens et al., 1999a). Therefore, only a limited number of independent transformants of each construct were taken for analysis. Transgenic T_2-seeds were analyzed for the presence of the modified Arc5a proteins using SDS-PAGE and Immunoblot analysis. For the Immunoblot analysis a rabbit polyclonal antiserum against the unmodified Arc5a protein was used. As can be seen in Figure 3.4, 3.6 and 3.7, all protein seed extracts from plants transformed with the modified *arc5-I* genes, regardless of the type of modification, reacted with the antibodies raised against unmodified Arc5a. The concentration of the modified Arc5a proteins was determined relative to the concentration of unmodified Arc5a found in seeds of the homozygous transgenic Colombia-O line, E105 (11.9 % of the total extractable seed protein; Goossens et al., 1999a). Protein extracts of the latter seeds were used as positive control in SDS-PAGE and Immunoblot analysis.

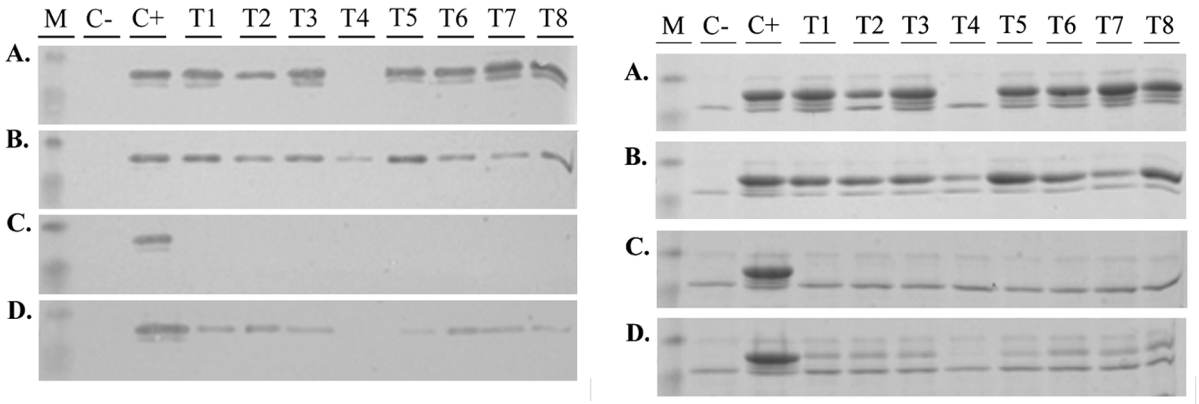


Figure 3.4: Immunoblot (left) and SDS-PAGE (right) analysis of crude protein extracts of *A. thaliana* seeds transformed with the *arc5I-M1* (A), *arc5I-M2* (B), *arc5I-M3* (C) and *arc5I-M4* (D) gene. Protein extracts were prepared of progeny (T_2) seeds of a non-transformed plant (C-) and plants transformed with the unmodified *arc5-I* gene (C+) and the modified genes (T1-T8). Lane M contains the marker proteins of which only the 36 kDa (upper) and 30 kDa (lower) bands are shown. The molecular masses of the unmodified and modified Arc5a proteins are approximately 32 kDa.

The modified Arc5a proteins harboring 3 or 4 substitutions with a methionine residue could be detected in transgenic *A. thaliana* seeds, except for the Arc5a-M3 protein. The first two modified Arc5a proteins tested were the Arc5a-M1 protein with a methionine at the position of Leu17, Leu41, Val44 and Leu47, and the Arc5a-M2 protein where the amino acids Ile58, Phe73 and Ile75 were substituted. Both modified proteins showed similar characteristics as the unmodified Arc5a and accumulated at high levels in the transgenic seeds (fig. 3.4A-B). The estimated accumulation levels in different lines varied from 7 to 17 % and from 3 to 13 % of the total extractable protein for Arc5a-M1 and Arc5a-M2, respectively. The modified Arc5a-M3 protein with substitutions at Leu91, Val94, Phe104 and Ile138, could not be isolated from

transgenic seeds (fig. 3.4C), although the transgene was detected in leaves of young transgenic T_2-seedlings through Southern analysis with a probe containing the coding sequence of the *arc5I-M3* gene (fig. 3.5). To exclude the possibility that the absence of the modified Arc5-M3 protein was unique to the individual plants analyzed, seeds from 14 additional independent transformants were analyzed using SDS-PAGE and Immunoblot analysis. Also in these transformants, the modified Arc5-M3 protein could not be detected (data not shown). Even when 10 µg instead of 1 µg protein was loaded for Immunoblot analysis, no modified Arc5a-M3 was detected. The Arc5a-M4 protein with substitutions at Phe148, Leu190, Val194 and Leu219 was observed when analyzing seed extracts, yet a much lower range of accumulation levels was seen on the Coomassie blue stained gel and the immunoblot than these of the first two modified Arc5a proteins, Arc5a-M1 and Arc5a-M2 (fig. 3.4D). The estimated accumulation levels were in the range of 1 to 4 % of the total extractable protein. Again, 7 more transformants were analyzed to exclude that the low expression levels were unique to the first analyzed transgenic seeds (data not shown).

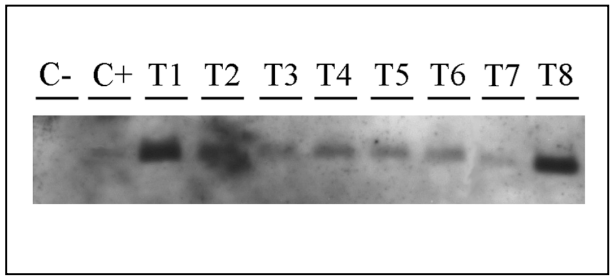


Figure 3.5: Southern analysis of non-transformed *A. thaliana* plants (C-) and *A. thaliana* plants transformed with the unmodified *arc5-I* gene (C+), and with the modified *arc5I-M3* gene (T1-T8). *SacI/XbaI*-digested DNA was probed with a *SacI/XbaI* fragment hybridizing to the coding sequence of the *arc5I-M3* gene in transgenic plants (see fig. 3.12). All transformants harbor at least one copy of the *arc5I-M3* gene.

The other two modified proteins contain a methionine-rich insertion in a variable loop of the Arc5a protein. Both modified proteins could be detected with SDS-PAGE and Immunoblot analysis, but the pattern of the bands for Arc5a was changed. This was the case for Arc5a-MI1 as well as Arc5a-MI2, but the pattern was different for both proteins (fig. 3.6-3.7).

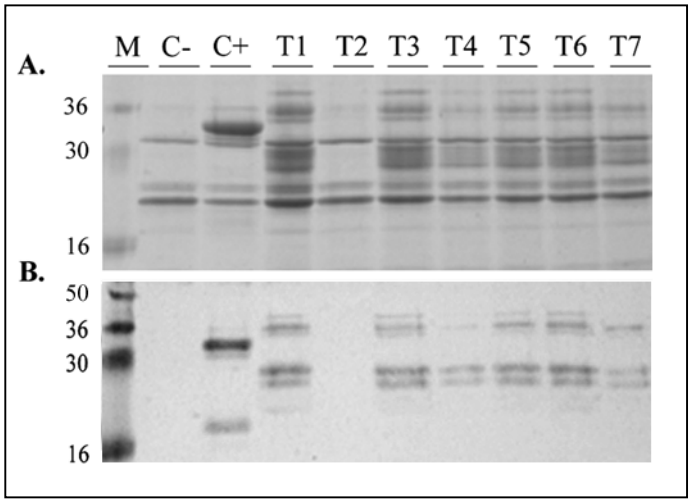


Figure 3.6: SDS-PAGE (A) and Immunoblot (B) analysis of crude protein extracts of *A. thaliana* seeds transformed with the *arc5I-MI1* gene. Protein extracts were prepared of progeny (T_2) seeds of a non-transformed plant (C-) and plants transformed with the unmodified *arc5-I* gene (C+) and the modified genes (T1-T7). Lane M contains the marker proteins of which the molecular mass is indicated on the left (kDa). The molecular mass of unmodified Arc5a is 32.2 kDa.

The expected banding pattern was one single band migrating around 33.2 kDa because, due to the insertion, the modified Arc5a-MI1 and Arc5a-MI2 proteins became approximately 1 kDa heavier than the unmodified Arc5a. At that position, only one weak band was detected for the Arc5a-MI1 protein. Additionally, four bands could be observed above and below the expected position for Arc5a-MI1 (fig. 3.6). The banding pattern for the Arc5a-MI2 protein deviated even more from the expected pattern and consisted of a number of closely spaced bands between 10 and 20 kDa (fig. 3.7). Most probably, these bands represent degradation products of the protein.

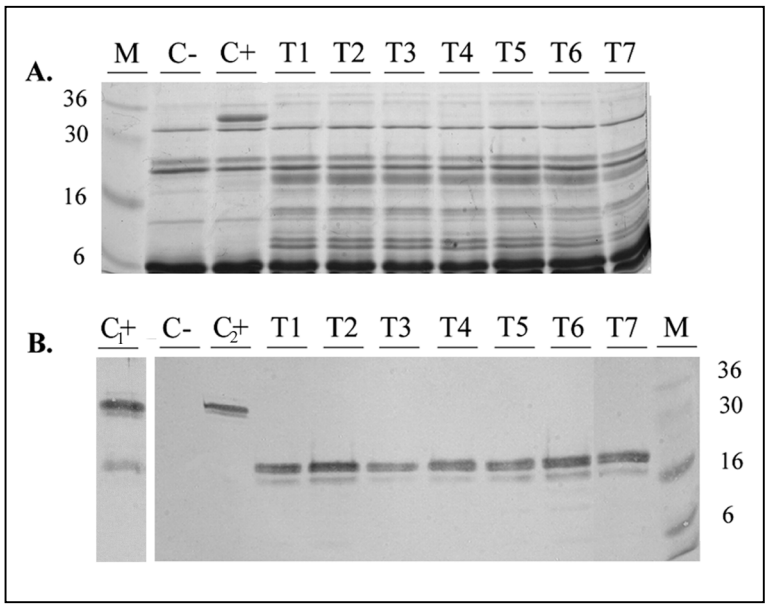


Figure 3.7: SDS-PAGE (A) and Immunoblot (B) analysis of crude protein extracts of *A. thaliana* seeds transformed with *arc5I-MI2*. Protein extracts were prepared of progeny (T_2) seeds of a non-transformed plant (C-) and plants transformed with the unmodified *arc5-I* gene (C+) and the modified genes (T1-T7). The protein extracts used for this immunoblot are prepared with the quick extraction procedure, except for C_1+, which was prepared with the normal extraction buffer without AEBSF. Lane M contains the marker proteins of which the molecular mass is indicated in kDa. The molecular mass of unmodified Arc5a is 32.2 kDa.

To be sure that degradation of the Arc5a-MI2 protein happened in seeds and not during extraction, a quick extraction procedure was used (see experimental procedures). The banding pattern for Arc5a-MI2 obtained after quick extraction (fig. 3.7) was identical to the pattern obtained with the normal extraction procedure (data not shown), indicating that the Arc5a-MI2 protein was degraded in seeds.

The accumulation level for these modified proteins ranged from 3 to 9 % and from 6 to 11 % of the total extractable protein for Arc5a-MI1 and Arc5a-MI2, respectively. The accumulation level of the Arc5a-MI2 is only a very rough estimation because it seems that not all degradation products could be detected with the antiserum (fig. 3.7, (A) vs. (B)).

Detection of the modified Arc5a proteins with combined groups of mutations

From the combined modified Arc5a proteins, only the Arc5a-M12 protein, combining the substitutions of Arc5a-M1 and Arc5a-M2, was found in transgenic seeds in a similar state as unmodified Arc5a and modified Arc5a-M1 and Arc5a-M2 proteins (fig. 3.8). Estimated accumulation levels were in the range of 4 to 14 % of the total

extractable protein. The other modified protein that could be isolated from transgenic *A. thaliana* seeds, although degraded, was Arc5a-M12I2 (fig. 3.9). The banding pattern seen on the Coomassie blue stained gel and on the immunoblot is situated between 10 and 20 kDa, similar to the banding pattern found for the Arc5a-MI2 protein. The accumulation level estimated on the basis of the detectable bands on the immunoblot varied from 1 to 3 % of the total extractable protein. This certainly is only a very rough estimation because not all degradation products could be detected with the antibodies (fig. 3.9, (A) vs. (B)).

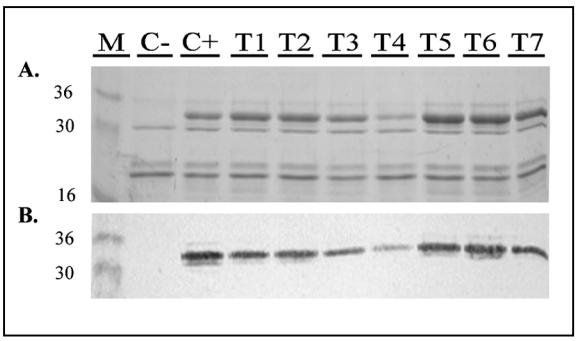


Figure 3.8: SDS-PAGE (A) and Immunoblot (B) analysis of crude protein extracts of *A. thaliana* seeds transformed with the *arc5I-M12* gene. Protein extracts were prepared of progeny (T_2) seeds of a non-transformed plant (C-) and plants transformed with the unmodified *arc5-I* gene (C+) and the modified genes (T1-T7). Lane M contains the marker proteins of which the molecular mass is indicated on the left (kDa).

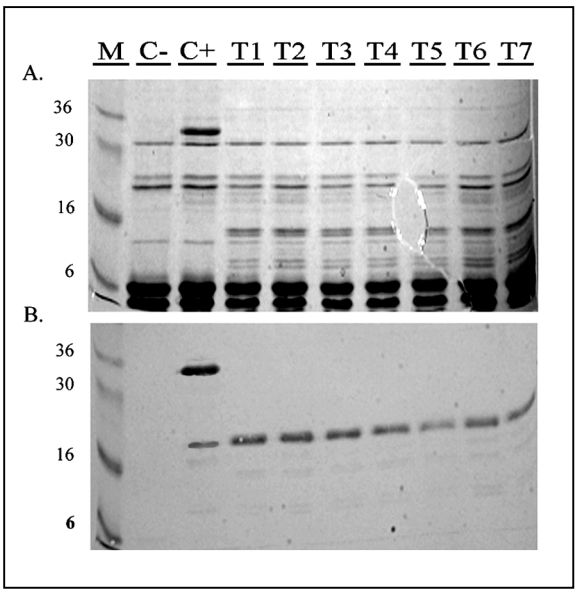


Figure 3.9: SDS-PAGE (A) and Immunoblot (B) analysis of crude protein extracts of *A. thaliana* seeds transformed with the *arc5I-M12I2* gene. Protein extracts were prepared of progeny (T_2) seeds of a non-transformed plant (C-) and plants transformed with the unmodified *arc5-I* gene (C+) and the modified genes (T1-T7). Lane M contains the marker proteins of which the molecular mass is indicated on the left (kDa).

All other combined modified Arc5a proteins (Arc5a-M12I1, Arc5a-M2I12 and Arc5a-M12I12) could not be detected on the Coomassie blue stained gel or the immunoblot under the used conditions. Since all the first modified proteins could be detected with the antibodies raised against unmodified Arc5a, there is little chance that the combined modified Arc5a proteins could not be detected on the immunoblot anymore, certainly because there were no proteins detectable on the Coomassie blue stained gel either.

Discussion

The seed storage protein arcelin-5a (Arc5a) from *Phaseolus vulgaris* was modified by creating substitutions with methionine codons and/or insertion(s) with a methionine-rich loop (see chapter 2). The stability of these modified Arc5a proteins was tested experimentally by expressing them in a seed-specific manner in *Arabidopsis thaliana*. This system was preferred over expressing the modified genes in *Escherichia coli* or yeast, or transiently expressing them, for instance in protoplasts from cell suspension cultures of bean (Giovinazzo et al., 1997). These latter methods are less time consuming, but don't provide complete information about the stability and accumulation of the modified proteins in seeds. A modified legumin *LeB4* gene from *Vicia faba* with 4 methionine codons and a glycigumin gene (hybrid gene constructed with the genes coding for glycinin of soybean and a modified legumin of *Vicia faba*) with 10 methionine codons could be produced in yeast as well as in a coupled *in vitro*-transcription/translation system. But when the same genes were introduced in tobacco, the corresponding proteins could not be detected in the seeds (Muntz et al., 1997).

Our final goal is to increase the methionine content of *Phaseolus* beans in particular and legumes in general. But the time needed to transform and regenerate *Phaseolus* plants, as well as most other legume plants is very long. Therefore, another dicotyledonous plant was chosen to test the stability of the modified Arc5a proteins. In seeds of dicotyledonous plants, the storage proteins are accumulated within specialized storage organelles, the protein storage vacuoles (PSV, Herman and Larkins, 1999). The unmodified Arc5a protein could be stored in the PSVs of *A. thaliana* at high levels showing that vacuolar targeting and storage of seed proteins in *Arabidopsis* seeds occurs very similar as in *Phaseolus* seeds (Goossens et al., 1999a). This characteristic, next to the its small size, short generation time and high seed production, made *A. thaliana* a good model plant to test the stability of the modified Arc5a proteins.

Three modified Arc5a proteins harboring three substitutions (Arc5a-M2), four substitutions (Arc5a-M1) or seven substitutions (Arc5a-M12) with a methionine could accumulate in transgenic *A. thaliana* seeds at very high levels, similar to unmodified Arc5a. This indicates that the stability of these proteins was not affected by the introduced modifications.

Three other modified Arc5a proteins with a methionine-rich insertion, Arc5a-MI1, Arc5a-MI2 and Arc5a-M12I2, could also be detected in transgenic *A. thaliana* seed extracts at high levels but with some clipping or degradation of the protein. For our goal, increasing the methionine content of the seeds, this is not dramatic since it is likely that all Arc5a protein products are still present in the plant seeds.

A possible explanation for the aberrant pattern seen on the Coomassie blue stained gel and immunoblot of the Arc5a-MI1 protein is that the glycosylation of the modified protein was altered. Before proteins can stably accumulate in the PSVs, they must pass through the secretory system which depends on the faithful execution of post-translational modifications. Inserting a methionine-rich loop can influence these modifications, one of which is glycosylation. The Arc5a protein has three glycosylation sites at Asn22, Asn70 and Asn79 and is co-translationally glycosylated in the ER with high-mannose glycans on two sites, one is certainly Asn22 (Goossens et al., 1994; Young et al., 1999). As the protein passes through the Golgi apparatus, the glycan chains are modified and fucosylated (Goossens et al., 1994). The protein conformation has been proven to play an important role in the modifications of glycans (Yet et al., 1988). Since the Asn22 glycosylation site is situated closely to the first insertion site, it is possible that the methionine-rich loop caused problems for glycosylation of the Arc5a-MI1 protein resulting in no glycosylation at this site giving a lighter protein. Another assumption for the aberrant pattern is that the insertion created a new cleavage site for plant enzymes resulting in a shorter protein without the N-terminal part. It's also possible that the methionine-rich loop itself possesses a cleavage site for plant enzymes since degradation of the protein was observed in all modified proteins containing a methionine-rich insertion at two different positions of the protein. The molecular masses of the protein products seen on the Coomassie blue stained gel and immunoblot can be in accordance with the estimated molecular masses of protein products obtained after cleavage within the insertion (approximately 29 and 4 kDa for Arc5a-I1 and 15 and 18 kDa for Arc5a-I2 and Arc5-M12I2).

It is clear that the methionine-rich insertion in the Arc5a-MI2 and Arc5aM12I2 protein had a big influence on the stability of the modified protein resulting in several small protein products. The upper bands situated around 18 kDa and 15 kDa can visualize protein products obtained after cleavage within the insertion, as described above. But these bands are also situated at the same place where in the extract of the positive control some breakdown products can be found (fig. 3.7B-3.9). In seeds transformed with the unmodified *arc5-I* gene, this breakdown process is for the most part caused by proteinases that are liberated during extraction since these breakdown products were not observed when an additional proteinase inhibitor mix (AEBSF) was added to the extraction buffer (data not shown). However, in most extractions of transformed seeds this AEBSF proteinase inhibitor mix was not used, resulting in this additional band around 18 kDa on the immunoblot. For the extracts of seeds transformed with the *arc5I-MI2* and *arc5I-M12I2* genes, these breakdown products are not produced by proteinases during the extraction because the same degradation pattern was found when the seed proteins were extracted with a quick extraction procedure (fig. 3.7B). These data suggest that the Arc5a protein harbors a 'protease sensitive' site. It is very likely that this site is closely situated to Asn111-Asn116 and that inserting a methionine-rich loop at this position has increased this sensitivity.

Amino acid sequence analysis of the different Arc5a related polypeptides will be needed to shed some light on possible cleavages or degradation that happened within the modified Arc5a-I1, Arc5a-I2 and Arc5a-M12I2 proteins.

The other modified protein with four substitutions, Arc5a-M4, showed much lower accumulation levels. Since 14 independent transgenic plant seeds were analyzed, attribution to position effect of the T-DNA insert may be rejected. One of the substitutions must have destabilized the structure of the protein (see chapter 2). The last modified protein with substitutions, Arc5a-M3, and the other modified proteins combining groups of mutations, Arc5a-M12I1, Arc5a-M2I12 and Arc5a-M12I12, could not be detected at all, pointing to an instable protein structure (see chapter 2). It's also possible that the accumulation level of these modified proteins was too low to detect. Under the used conditions, the detection limit of the Immunoblot analysis was approximately 0.5 % Arc5a. Anyhow, such low accumulation levels are not sufficient to enhance the methionine content of seeds.

The above results clearly show that the Arc5a protein can tolerate a significant amount of codon changes (up to seven) without changing the stability of the protein. The resulting increase of methionine content of the Arc5a protein (from 0 to 7 methionines of the 240 amino acids) is the best improvement published till now. It should be possible to increase this number of methionine replacements using the least risk-bearing substitutions from group 3 and group 4 like Val94, Ile138 and Leu219.

Experimental procedures

Plant material

Arabidopsis thaliana land race Colombia-O seeds (Lehle Seeds, Round Rock, USA) were sown in soil (with ¼ vermiculite, ½ potting soil and ¼ slipping soil) under a 16 h light/8 h dark regime at 22 °C in the greenhouse. Seedlings were separated and each transferred to a 5 cm lightweight plastic pot. The first bolts were clipped to encourage proliferation of many secondary bolts. When plants were flowering and the first siliques were formed (fig. 3.10), they were used for transformation. After transformation, plants were kept in the greenhouse until maturity and seeds were formed.

Figure 3.10: Developmental stage of the plants used for transformation.

Transformed T_1-plants were obtained from surface sterilized seeds grown *in vitro* in 100x20 mm Falcon® Petri dishes (Sigma) on growth medium (Valvekens et al., 1988) supplemented with 50 µg/ml kanamycin and 200 µg/ml cefotaxime. After selection (fig. 3.11), seedlings were transferred to soil and grown in the greenhouse. Seeds of at least 25 plants were harvested and used for analysis. These T_2-seeds are a mixture of homozygous transformed, homozygous non-transformed and hemizygous seeds.

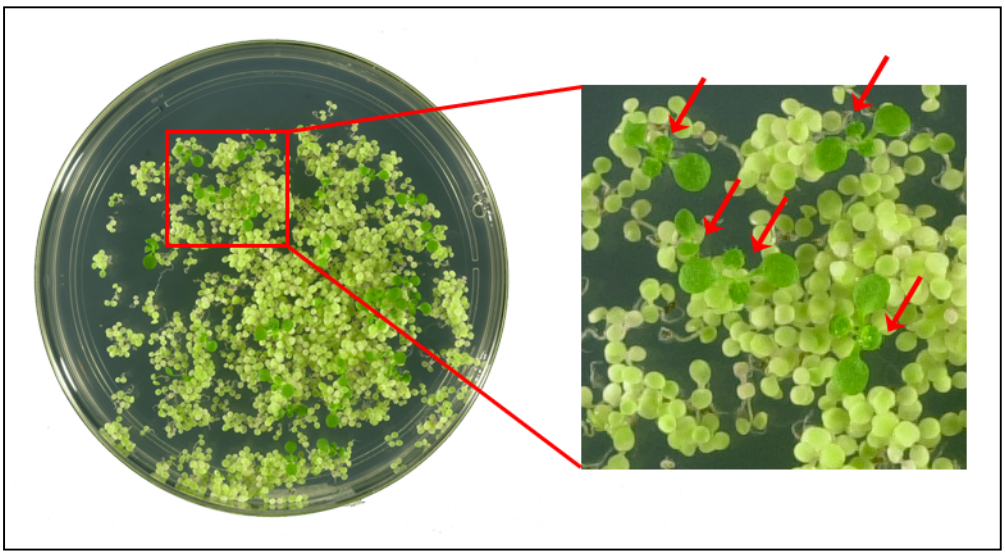


Figure 3.11: Selection of transformed T_1-plants on growth medium with 50 mg/l kanamycin. The red arrows indicate the transformed seedlings between the non-transformed ones.

Bacterial strains and plasmids

Agrobacterium tumefaciens strain C58C1Rif^R (Holsters et al., 1980) containing pMP90 (Koncz and Schell, 1986), a nopaline-type non-oncogenic Ti plasmid, and a binary vector derived from pATAG3 (Goossens et al., 1999a) was used for transformation. All the binary vectors contain between the T-DNA borders the two marker genes neomycin phosphotransferase II (*nptII*) and the *Escherichia coli* β-glucuronidase (*uidA*) gene (Jefferson, 1987) with the potato *st-ls1* intron (Vancanneyt et al., 1990). In addition, they harbor the genomic *EcoRI* fragment containing the coding sequence for the unmodified or for one of the modified *arc5-I* genes: pATARC3-B1b contains the unmodified *arc5-I* gene (Goossens et al., 1999a), pATARC3-M1 contains the modified *arc5I-M1* gene, pATARC3-M2 the modified *arc5I-M2* gene, … (fig. 3.3). This *EcoRI* fragment represents the largest *arc5-I* sequence available from the genomic clone (Goossens et al., 1995), harboring potential TATA and CCAT boxes, the majority of the potential *cis*-regulatory elements for seed-specific expression and three potential 5' MAR sequences (see chapter 1).

Agrobacterium tumefaciens transformation

Transformed *A. thaliana* plants harboring the modified *arc5I-M1*, *arc5I-M2*, *arc5I-M3*, *arc5I-M4*, *arc5I-MI1* and *arc5I-MI2* genes were obtained using the vacuum infiltration method as described by Bechtold and co-workers (1993). Flowering plants from the greenhouse (fig. 3.10) were pulled out of the soil and collected in a glass jar (14.5 cm height). The pots containing plants to be infiltrated were soaked into an *Agrobacterium* solution so that the entire plants were covered, including the rosette, and were placed under a vacuum of 400 mm Hg for 15 min. After a quick release, the plants were washed with water. The *Agrobacterium* solution was obtained from a 500 ml culture grown for 24 h at 28 °C on a gyratory shaker (yeast extract broth (YEB)(Grimsley et al., 1986) medium supplemented with the appropriate antibiotics, inoculated with a 3 ml overnight starter culture). Cells were collected by centrifugation and washed in 100 ml infiltration medium (2.15 g MS salts, 1 ml B5 vitamins (Sigma-Aldrich, Irvine, UK), 50 g sucrose, 0.044 µM benzylaminopurine (BA, Sigma-Aldrich), 200 µl Silwet L-77 (Lehle Seeds), pH 5.6) and resuspended to a final OD_{600} of 1.

Transformed *A. thaliana* plants harboring the modified *arc5I-M12*, *arc5I-M12I1*, *arc5I-M12I2*, *arc5I-M2I12* and *arc5I-M12I12* genes were obtained using the floral dip method (Clough and Bent, 1998). Here, the flowering plants didn't have to be pulled out of the soil because only the flowers were dipped into the A*grobacterium* suspension. The dipping lasted for 10 seconds whereupon the plants were put in the dark for 24 h at 22 °C without washing. The *Agrobacterium* solution was obtained from a 500 ml culture grown for 24 h at 28 °C on a gyratory shaker (YEB medium supplemented with the appropriate antibiotics, inoculated with a 3 ml overnight starter culture). Cells were collected by centrifugation and resuspended in a 5 % sucrose-0.05

% Silwet L-77 solution to a final OD_{600} of 0.8. After 7 days, the dipping was repeated under exactly the same conditions.

DNA analysis of transgenic plants

Genomic DNA was prepared from leaves of plants transformed with the *arc5I-M3* gene using the CTAB method as described by Dellaporta and co-workers (1983). DNA (5 μg) was digested with *Sac*I-*Xba*I, fractionated on a 1 % agarose gel by electrophoresis, transferred to a Hybond-N^+ membrane (Amersham Pharmacia Biotech) and hybridized at 60 °C with a 1.1-kb *Sac*I/*Xba*I *arc5I-M3* fragment (fig. 3.12). The probe was labeled with fluorescein-dUTP by using 'the Gene Image Random Prime Labeling Module' kit (Amersham Pharmacia Biotech). Detection was performed using 'the Gene Images CDP-Star Detection Module' system (Amersham Pharmacia Biotech) with the help of alkaline phosphatase. All handlings were performed as prescribed by the manufacturer's guidelines.

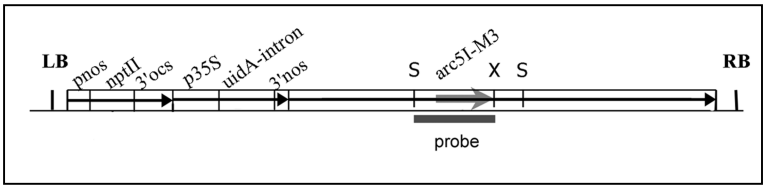


Figure 3.12: The transferred region of the binary vector pATARC3-M3 with indication of the coding sequence for the *arc5I-M3* gene (arrow), the probe used in Southern analysis to detect the presence of the *arc5I-M3* gene (bar) and the restriction sites *SacI* (S) and *XbaI* (X).

Protein analysis of transgenic plants

Crude seed protein extracts were obtained as described by Goossens et al. (1999a). Grounded seeds were extracted twice with hexane to remove lipids. The residue was lyophilized and subsequently extracted twice with extraction buffer (50 mM NaCl, 50 mM Gly, protease inhibitor mix (1 tablet/10ml, 2× CØmplete (Roche Diagnostics, Brussels)), pH 2.4) for 15 min at room temperature under continuous shaking. For the first extractions, a second protease inhibitor AEBSF (4-(2-aminoethyl)benzene-sulfonyl fluoride; ICN, California, US) was added to the extraction buffer. Total protein quantity in the crude extracts was determined by the Lowry method using the DC protein assay (Bio-Rad, Hercules, California, US) with Bovine serum albumin (BSA, Sigma-Aldrich) as a standard. 20 μg proteins were separated on a 12.5 % polyacrylamide gel using SDS-PAGE and visualized by Coomassie Blue staining. Immunoblot analysis of 1 μg of proteins was performed using a rabbit polyclonal antiserum against unmodified arcelin-5 after transferring the proteins to a membrane (Problott™ Membranes, Applied Biosystems, Foster City, California). Estimations of the protein accumulation levels were conducted by analyzing a digitalized grey-scale image of the immunoblot using the Imagemaster® VDS Software (Amersham Pharmacia Biotech, Aylesbury,

UK). The concentration of the modified proteins was determined relative to the concentration of the positive control (seeds of a homozygous transgenic Colombia-O plant, accumulating 11,9 ± 0,1 % unmodified Arc5a, Goossens et al., 1999a).

A quick extraction procedure was used to test whether degradation of the modified Arc5a-MI2 protein happened during the normal extraction procedure. Therefore, 10 mg of seeds were crushed in liquid nitrogen, immediately resolved in 75 µl sample buffer (2 ml 1M Tris.HCl pH 6.8, 3.2 ml glycerol, 2.6 ml H_2O, 1.6 ml β-mercapto-ethanol, 0.6 ml 10 % SDS) and denatured at 95 °C for 5 min. The pellet was collected by centrifugation (13000 tpm, 4 min) and 1 ml 90 % cold aceton was added for 30 min at –20 °C. After centrifugation (8000 tpm, 4 min) the pellet was washed with 1 ml 75 % ethanol and resolved in 500 µl sample buffer (+ bromophenol blue). After a second denaturation, the proteins were analyzed immediately.

TRANSFORMATION OF PHASEOLUS BEANS

Introduction

The genus *Phaseolus*

The *Phaseolus* beans are a member of the family of the *Fabaceae* (*Leguminosae*), which comprises about 600 genera and about 13000 species. The genus *Phaseolus* includes approximately 200 species of plants, many of which are cultivated as food or garden ornamentals. The five cultivated species are *P. vulgaris*, *P. acutifolius*, *P. lunatus*, *P. coccineus* and *P. polyanthus.* Most of the dry bean production is confined to *P. vulgaris* (common bean). This species is cultured in over 70 countries in temperate, sub-tropical and tropical climates worldwide. The major producers are Brazil, China, United States and Mexico (Singh, 1999). Together they produce more than 20 million ton per year (Parker, 1995; Singh, 1999). The name *P. vulgaris* refers to hundreds of varieties and cultivars of the common bean which have been in cultivation for thousands of years (fig. 4.1). Beans provide energy and proteins for millions of people, mostly with low incomes, in Central and South America, Africa and India.

Figure 4.1: Some varieties of the common bean (*P. vulgaris*).

P. acutifolius (tepary bean) is nutritionally very similar to *P. vulgaris*, but unlike it, is very well adapted to environmental or pathological stresses. *P. acutifolius* plants (fig. 4.2) can grow in arid zones and are resistant to microbiological invasion and to many beetles which attack other beans (Sotelo, 1996). These resistance traits (reviewed in Pratt and Nabhan, 1988) are not found *in P. vulgaris* but can be introduced in the latter through interspecific hybridization. Introgression of a simple inherited trait from *P. acutifolius* to *P. vulgaris* is feasible as has been demonstrated for *Xanthomonas campestris* resistance (Scott and Michaels, 1992). Tepary beans are mainly found in Central and South-west America and in the North and South-east of Mexico and Guatemala (Sotelo, 1996). It is difficult to estimate even roughly the production of *P. acutifolius* beans because most are used for local consumption. Although people who

eat them find the flavor excellent, this species is not widespread. Compared with the common bean, there is less seed variability although seeds exist in many colors from white, grey, yellow, purple, and coffee to black.

Figure 4.2: Flowering *P. acutifolius* plants with maturing pods.

Regeneration of *Phaseolus* beans

The main problem when establishing a transformation procedure for grain legumes is their recalcitrance to *de novo* regeneration. For *Phaseolus* beans, most of the available regeneration procedures have in common that shoots are produced from embryo or seedling explants without a phase of unorganized growth (a callus phase). Shoots proliferate in a region adjacent to meristems or originate from pre-existing meristems that are contained in the explants. The explants used in these regeneration experiments contained seedling apical meristems (Kartha et al., 1981; Martins and Sondahl, 1984; Allavena and Rossetti, 1986; Malik and Saxena, 1992) or meristems located in the cotyledon axils (McClean and Grafton, 1989; Franklin et al., 1991; Malik and Saxena, 1992; Mohammed et al., 1992a; 1992b; Vaquero et al., 1993, Cruz de Carvalho et al., 2000). Till now, such a regeneration system could not be combined with *Agrobacterium*-mediated gene transfer. Most probably, the meristematic cells from which plants regenerate are not exposed to *Agrobacterium* during co-cultivation.

Another problem with these regeneration procedures is that regenerating shoots originate from multiple meristematic cells and will therefore be chimeras (transformed and non-transformed parts).

A regeneration procedure involving a callus phase may be more suitable to be combined with *Agrobacterium* in order to recover transgenic plants. The rapidly dividing cells in the initial explant or in the proliferating callus itself are likely targets for *Agrobacterium*. Moreover, a callus phase makes rigorous selection for transformed tissues and regeneration from such tissues feasible. Mohamed and coworkers (1993) published the first report about regeneration from callus in *P. vulgaris*. Green nodular callus was obtained from pedicel explants of the breeding lines Xan-159 and GN Tara. Strikingly, the two genotypes for which this type of generation was achieved had been constructed from interspecific crosses of *P. vulgaris* and *P. acutifolius*, suggesting that their competence for regeneration derives from the latter species. In a screening of a range of *P. acutifolius* genotypes, regeneration capacity was indeed encountered (Dillen et al., 1996). Green nodular callus with a competence to regenerate was established from floral or vegetative buds or from pedicels of *P. acutifolius* using a combination of thidiazuron and indole-3-acetic acid (fig. 4.3). Of twelve genotypes tested, seven were responsive in that they produced morphologically normal shoots. Of six genotypes (NI574, NI576, NI589, NI787, NI845 and NI1299), fertile regenerants were established in the greenhouse. The regeneration capacity occurred in wild as well as in cultivated genotypes. On the other hand, screening of a wide range of *P. vulgaris* accessions revealed little regeneration potential (Zambre et al., 1998b). In subsequent experiments, regeneration-competent callus of the *P. vulgaris* breeding line Xan-159 and of the *P. acutifolius* cultivars NI574, TB1 and PI440795 could be obtained from mature embryo explants (Zambre et al., 1998a).

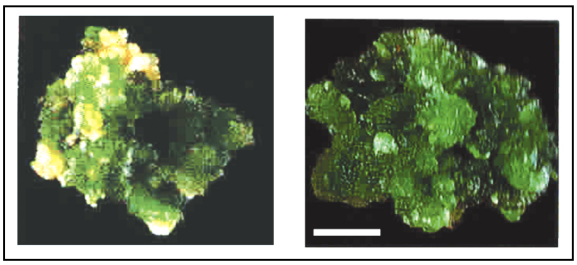


Figure 4.3: Green nodular callus established from vegetative buds of the *P. acutifolius* wild genotype NI576 (left) and from cotyledons of the *P. acutifolius* cultivar TB1 (right). Bar = 6 mm

Transformation of *Phaseolus* beans

With higher plants, in principle two different gene delivery systems can be applied: indirect *Agrobacterium*-mediated transformation or direct transformation of protoplasts, cells or tissues. Because of the lack of *in vitro* regeneration capacity in the genus *Phaseolus,* investigators have been prompted to devise 'regeneration-independent' approaches. Russell and colleagues (1993) obtained the first transgenic *P. vulgaris* plant using such an approach. Seedling apical meristems were submitted to particle bombardement and after stimulation of axillary shoot proliferation, chimeric

plants were obtained. Putative germline transformants were identified by histological *β-glucuronidase* assays. Plants that passed the transgenes on to the progeny were recovered at a frequency of 0.03 % of the harvested shoots. The transgenic plants contained multiple and sometimes incomplete copies inserted at one locus. The transgenes were stably inherited through many generations. The method was successful for one cultivar of *P. vulgaris* out of four tested. Using a similar strategy, Aragão and colleagues (1996, 1999) obtained germline transformants of another cultivar of *P. vulgaris*. Again, relatively low transformation frequencies were obtained.

For introduction of foreign genes into higher plants, *Agrobacterium tumefaciens*-mediated transformation is generally preferred to direct gene delivery methods, such as particle bombardment, because the former method offers several advantages, e.g. integration of a well defined DNA sequence, potentially low copy number, high co-expression of the introduced genes, and preferential integration into active regions of the chromosome (Birch, 1997; Gheysen et al., 1998).

In our opinion, the regeneration system established for *P. acutifolius* involving a callus phase (Dillen et al., 1996) should be compatible with *Agrobacterium*-based transformation. Therefore, several genotypes were tested for transformation-capacity resulting in one transgenic line for the genotype NI576 (Dillen et al., 1997a). Next, the influence of different co-cultivation conditions on the efficiency of *Agrobacterium*-mediated gene transfer was assessed to improve the transformation efficiency (Dillen et al., 1997b; De Clercq et al., 2002). Application of the optimized transformation conditions resulted in a reproducible transformation procedure for *P. acutifolius* capable of producing ten independent transgenic lines in one transformation experiment (De Clercq et al., 2002).

Agrobacterium-mediated transformation of *Phaseolus acutifolius*

abstracted from Dillen W., De Clercq J., Goossens A., Van Montagu M., Angenon G. (1997a). *Agrobacterium*-mediated transformation of *Phaseolus acutifolius*. Theor Appl Genet 94, 151-158.

Results

Regeneration-competent callus established from bud explants (Dillen et al., 1996), was co-cultivated five days after the third subculture with the *Agrobacterium tumefaciens* strain C58C1Rif^{R}(pMP90) harboring a binary vector with the neomycin phosphotransferase II (*nptII*) and β-glucuronidase (*uidA*) marker genes. After co-cultivation explants were put on CIM1 with 500 mg/l cefotaxime and 20 mg/l geneticin for selection. In five out of six tested genotypes, transgenic *uidA* expression was detected 4 days after co-cultivation (fig. 4.4a).

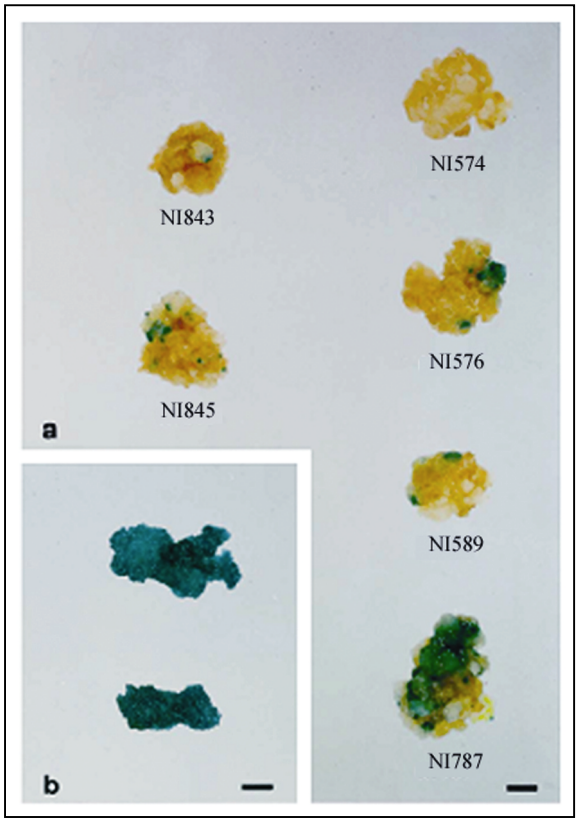


Figure 4.4: GUS activity in callus of different *P. acutifolius* genotypes 4 days after co-cultivation with *Agrobacterium* (a) and geneticin-resistant callus of NI843 (b). Bar = 1.2 mm (derived from Dillen et al., 1997a)

The recalcitrance of the sixth genotype, NI574 for transformation was confirmed in other experiments involving more plant material. At the end of the third subculture of co-cultivated material on geneticin-containing CIM1, thirteen callus lines of genotype NI576, two of NI787, and seven of NI843 had survived selection (fig. 4.4b). Callus from one line, genotype NI 576, a wild genotype identified as possessing superior regeneration capacity (Dillen et al., 1996), continued to proliferate under selective conditions and produced viable GUS-positive shoots (fig. 4.5) which could be grafted and hardened in the greenhouse. This line had been transformed with a binary plasmid which, in addition to the marker genes, contained a genomic fragment encoding the unmodified arcelin-5a protein. Mendelian segregation in the offspring for GUS activity and geneticin resistance together with the occurrence of the Arc5a protein, constitute proof for stable transformation and transmission of the transgenes. The concentration of Arc5a was in the range of 20-25 % of total seed protein (Dillen et al., 1997a; Goossens et al., 1999a).

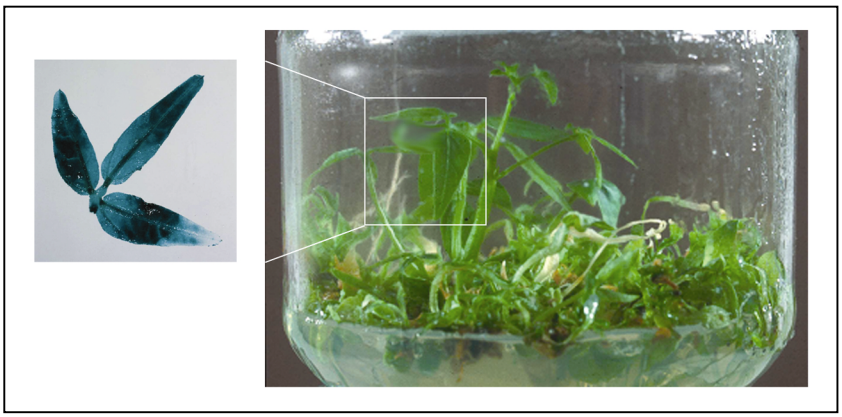


Figure 4.5: GUS activity in a leaf of an *in-vitro* formed transgenic shoot of NI576.

Experimental procedures

Establishment of green nodular callus

P. acutifolius plants of several genotypes (NI = collection number of the National Botanic Garden of Belgium, Meise, Belgium) were grown in a greenhouse under 8 h light/16 h dark regime (light intensity 50 µmol/m^2/s) at 24 °C. Bud explants were excised 1 month after sowing. The apical and axillary bud clusters were dissected and separated into individual buds of approximately 1 mm (fig. 4.6) and cultured on callus induction medium (CIM1; Murashige and Skoog (1962) salts and organic addenda, 20 g/l sucrose, 8 g/l bacto-agar (Difco Labs, Detroit, Michigan, USA), 0.25 mg/l indole-3-acetic acid, 0.5 mg/l thidiazuron, pH 5.7) to induce green nodular callus. Initial cultures were kept at 24 °C in the dark for 1 week. Thereafter, cultures were transferred to a 16-h light/ 8-h dark regime (light intensity 20 µmol/m^2/s) at 24 °C. After three cultures lasting for 2, 1 and 1 months, respectively, 2- to 4-mm callus pieces were used for co-cultivation. On average, five bud explants per genotype were used to obtain 20 of these callus pieces.

Figure 4.6: Apical and axillary buds from a one month old *P. acutifolius* NI576 plant used as explants for callus proliferation.

Co-cultivation of callus

Callus was pre-cultured for 5 days on CIM1, chopped into 2- to 4-mm pieces, and cultured in 100x20 mm Falcon® Petri dishes (Sigma) sealed with gas-porous tape (Urgopore, Chenoves, France) at a density of 20 pieces per 32 ml of co-cultivation medium in the presence of 20 µM of 3',5'-dimethoxy-4'-hydroxy-aceto-phenone (acetosyringone). The co-cultivation medium was as CIM1 but without agar and including 10 mM of D(+)-glucose and 20 mM of 2-[N-morpholino] ethane sulphonic acid (MES), pH 5.5. *A. tumefaciens* inocula were obtained from cultures grown for 36 h on a gyratory shaker (150 rpm) at 28 °C. Cells (100 ml) were collected by centrifugation (2000 *g*, 10 min), washed twice in 50 ml co-cultivation medium, and

pre-incubated for 4 h at room temperature in 5 ml co-cultivation medium with 200 µM acetosyringone. The concentrated *Agrobacterium* suspension was added to the explants to a final OD_{600} of 0.8. The temperature was 24 °C with a light period of 16 h (light intensity 20 µmol/m^2/s). After 2 days, explants were washed twice in co-cultivation medium containing 500 mg/l cefotaxime, blotted dry, and cultured on selective medium (CIM1 with 20 mg/l geneticin).

Bacterial strains and plasmids

Co-cultivation was carried out with the *A. tumefaciens* strain C58C1RifR (Holsters et al., 1980) containing the non-oncogenic plasmid pMP90 and the binary plasmid pATARC3-B1b which contains between the T-DNA borders the two marker genes *nptII* and *uidA*-intron, as well as a genomic fragment coding for the unmodified Arc5a protein (see chapter 3, fig 3.3).

Selection and regeneration

All media used for selection and regeneration of shoots contained 20 mg/l geneticin (antibiotic G418). After washing, the callus was transferred to CIM1 supplemented with 500 mg/l cefotaxime. After 14 days, surviving tissue was subcultured at 3-week intervals on the same medium. The cefotaxime concentration was reduced to 400 mg/l in the first subculture, to 300 mg/l in the second subculture, to 200 mg/l in the third subculture, and cefotaxime was omitted in the fourth and subsequent subcultures. In the fifth subculture the callus was transferred to shoot induction medium (SIM; MS salts and organic addenda, 20 g/l sucrose, 8 g/l bacto-agar, 10 % coconut water (Sigma, St. Louis, Missouri, USA), 1.0 mg/l benzyl amino purine (BA), pH 5.7) for 1 month followed by two 2-week passages to shoot development medium (SDM; as SIM but with 0.1 mg/l BA). Shoots were excised from the last culture on SDM and grafted. Established grafts were transferred to potting soil in the greenhouse and covered with plastic foil. The relative humidity was reduced by gradually removing the foil. Hardened plants were grown in the greenhouse for 1 month and then transferred to a growth chamber and provided with short-day conditions (8 h light, 50 µmol/m^2/s, 24 °C) to induce flowering. Seeds were harvested between 2 and 3 months after transfer to the greenhouse.

Gus assay

GUS activity was histochemically localized according to Jefferson (1987). Tissue was incubated in staining buffer (100 mM $NaPO_4$, 50 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid cyclohexyl-ammonium salt, 0.1 % β-mercaptoethanol, 0.1 % Triton X-100, pH 7.2) for 16 h. Stained plant material was cleared with 70 % ethanol.

Optimization of the transformation procedure for *P. acutifolius*

adapted from De Clercq J., Zambre M., Van Montagu M., Dillen W., Angenon G. (2002). An optimized *Agrobacterium*-mediated transformation procedure for *Phaseolus acutifolius* A. Gray. Plant Cell Rep, in press.

Results and discussion

The basic transformation procedure for *P. acutifolius* was as described above. Regeneration-competent callus of the genotype NI576, established from bud explants (Dillen et al., 1996), was co-cultivated 5 days after the third subculture. The *Agrobacterium* inocula were pre-incubated with 200 µM AS for 4 h and then added to the callus explants at an initial OD_{600} of 0.8. Co-cultivation was performed for 48 h in liquid co-cultivation medium containing 10 mM glucose, 20 µM AS, and 20 mM MES, pH 5.5. After co-cultivation, the calli were washed and cultured on non-selective medium. The influence of changing the transformation conditions was assessed by using *Agrobacterium* strains with a binary vector harboring the *uidA*-coding sequence interrupted by an intron (pTJK136 or pATARC3-B1b). Transient expression of the *uidA*-intron gene was determined histochemically 4 days after co-cultivation of the calli.

In a first comparative experiment, calli were infected with the *Agrobacterium* strain C58C1Rif^R(pMP90)(pTJK136) collected from either early-log phase (12 h growth) or late-log phase (36 h growth) cultures. The bacteria were added at the same optical density (OD_{600} = 0.8). Infection with bacteria from an early-log phase yielded much more *uidA* expression zones than infection with bacteria in a late-log phase (fig. 4.7A; table 4.1). This experiment was repeated three times and always gave the same result.

growth phase	experiment 1		experiment 2	
	large	small	large	small
12h growth	14	28	11	33
36h growth	2	15	3	18

Table 4.1: Number of large and small *uidA* expression zones observed in two representative experiments with 30 explants each. The calli were co-cultivated with bacteria of the strain C58C1Rif^R(pMP90)(pTJK136) harvested in early-log phase (12 h growth) or in late-log phase (36 h growth).

In contrast to our results, tests performed by De Bondt and co-workers (1994) on transformation of apple leaf explants with three different strains in either lag phase, mid-log phase or stationary phase showed no significant effect of the bacterial growth phase. Because of the impact of the bacterial growth phase on the infection of *P. acutifolius* calli, all subsequent experiments were performed with early-log phase inocula.

To determine the influence of temperature during co-cultivation, experiments were performed with temperatures in the range of 15 °C to 29 °C. As described by Dillen and colleagues (1997b), the optimum temperature for co-cultivation was 22 °C, irrespective of the type of non-oncogenic Ti plasmid. *UidA* expression markedly decreased when the temperature was increased to 25 °C. Very low levels of *uidA* expression resulted at 27 °C and 19 °C and no expression was observed at 29 °C and 15 °C. The difference in number of *uidA* expression zones at 22 °C and 25 °C is notable (fig. 4.7B; table 4.2). Similar results were obtained when the effect of temperature was investigated in garlic transformation (Kondo et al., 2000). The highest transient *uidA* expression in garlic calli was observed at 22 °C whereas the ratio of GUS-stained calli to total calli decreased by 85 % at 20 °C and by 69 % at 24 °C. Also results on stable transformation of cotton were in line with our experiments (Sunilkumar and Rathore, 2001). Co-cultivation of cotyledon disks at 21 °C, compared to 25 °C, consistently resulted in higher transformation frequencies.

temperature / plasmid	25 °C				22 °C			
	experiment 1		experiment 2		experiment 3		experiment 4	
	large	small	large	small	large	small	large	small
pMP90	14	28	11	33	29	38	32	44
pEHA101	3	19	0	8	2	15	3	18
pGV2260	2	24	5	24	14	30	30	36

Table 4.2: Number of large and small *uidA* expression zones in four experiments with 30 callus pieces each. The calli were co-cultivated at two different temperatures with bacteria of the strain C58C1Rif^R(pTJK136) and three different non-oncogenic Ti plasmids.

To examine the effect of the type of non-oncogenic Ti plasmid, a nopaline-type plasmid (pMP90) was compared with an octopine-type (pGV2260) and a succinamopine/ agropine-type (pEHA101) together with the binary vector pTJK136. From experiments performed at 25 °C, it was obvious that the efficiency of DNA transfer reached the highest level with the nopaline helper plasmid. The number of *uidA*-expressing zones seen with pMP90 is much higher than that with pGV2260 and pEHA101 (fig. 4.7B, upper rows; table 4.2, experiments 1 and 2). Also at 22 °C, the highest GUS activity was obtained with the strain carrying pMP90, although the difference with the strain carrying pGV2260 was not as marked as at 25 °C (fig. 4.7B, lower rows and table 4.2, experiments 3 and 4). As a result, we decided to use only pMP90 as helper plasmid in the subsequent experiments. All experiments clearly showed that pEHA101 was the least efficient helper plasmid for *P. acutifolius* callus transformation. This observation is

surprising because the *Agrobacterium* strain EHA101 (or its derivative EHA105, or its oncogenic progenitor A281) is generally recognized as superior in facilitating gene transfer to plant cells, e.g., in leguminous species, such as soybean (Meurer et al., 1998; Donaldson and Simmonds, 2000), pea (Nadolska-Orczyk and Orczyk, 2000), and peanut (Egnin et al., 1998), but also in other crops, such as apple (De Bondt et al., 1994), cabbage (Takasaki et al., 1997), blueberry (Cao et al., 1998), and wheat and barley (Guo et al., 1998).

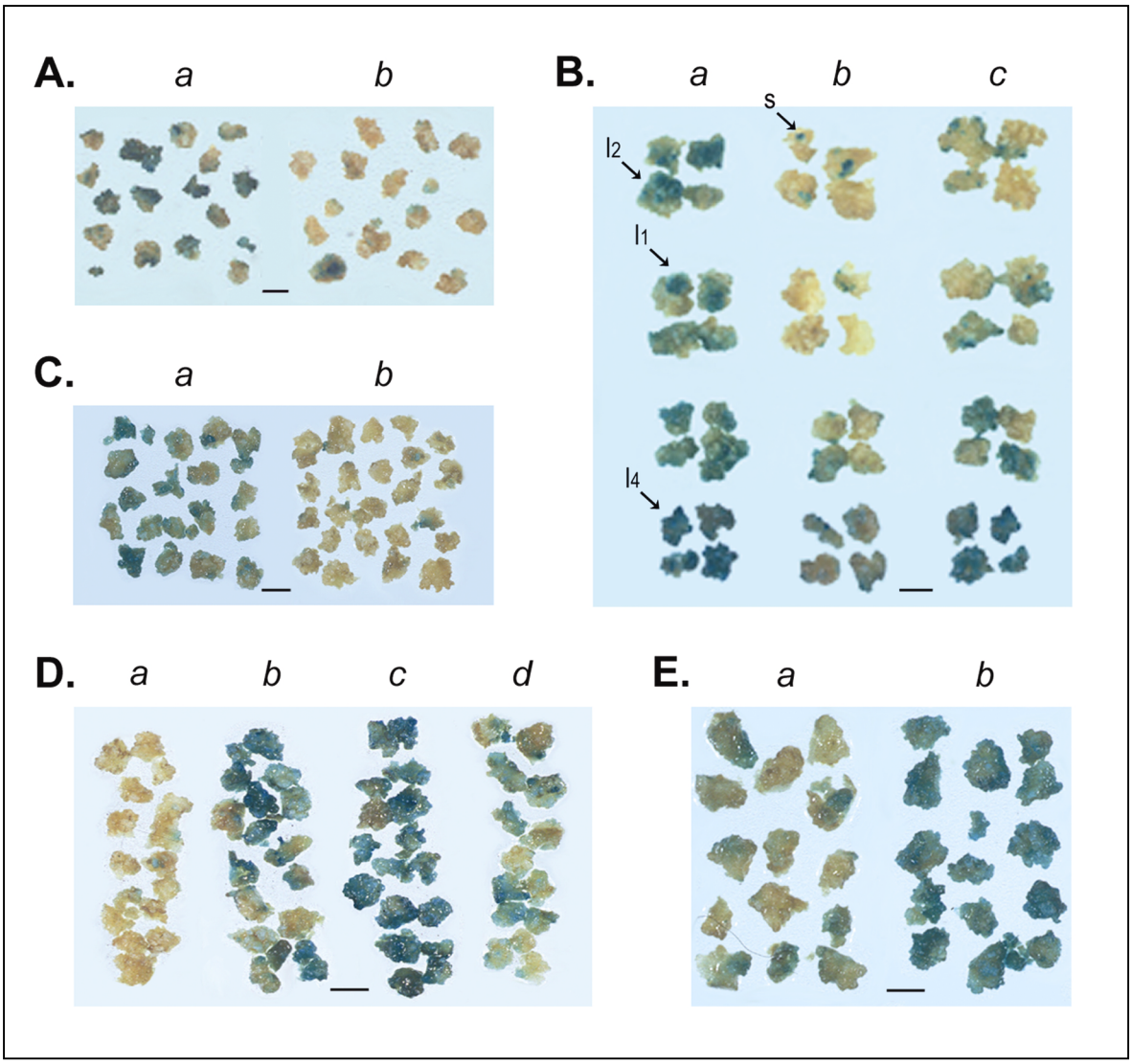


Figure 4.7: Histological assay of *uidA* expression after co-cultivation with (**A)** bacteria of the strain C58C1Rif^R(pMP90)(pTJK136) harvested in early-log phase (*a*) and in a late-log phase (b); (**B)** bacteria of the strain C58C1Rif^R (pMP90)(pTJK136) (*a*), EHA101(pTJK136) (*b*), and C58C1Rif^R(pGV2260)(pTJK136) (*c*) at 25 °C (upper two rows) and 22 °C (lower two rows) (only the four highest expressing calli are shown); (**C)** bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) under a 16-h light/8-h dark photoperiod (*a*) or complete darkness (*b*); (**D)** bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) in the presence of 0, 20, 200 and 2000 µM AS (*a-d*); (**E)** bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) in co-cultivation medium without (*a*) and with (*b*) buffering with 20 mM MES. To evaluate the transformation frequency a distinction was made between small spots (s) and large spots (l). Each *uidA* expression zone covering about a quarter of the explant was counted as a large spot: l_1=1 large spot; l_2=2 large spots; l_4=4 large spots. Bar = 3 mm

However, the fact that this strain is not better than other strains, or even worse, has also been observed in *Phaseolus vulgaris* (Zhang et al., 1997) and in other plants, such as cotton (Sunilkumar and Rathore, 2001), rice (Hiei et al., 1994; 1997), kalanchoe (Jia et al., 1989), black locust (Igasaki et al., 2000) and hibiscus (Srivatanakul et al., 2001). This underscores the importance of testing various *A. tumefaciens* strains for every species and genotype under study.

To test the effect of the light conditions during co-cultivation on gene transfer, calli were co-cultivated in complete darkness or under a 16-h light/8-h dark cycle. In darkness, almost no GUS activity could be detected (fig 4.7C; table 4.3). Moreover, the explant survival rate was drastically reduced. We concluded that co-cultivation in darkness was deleterious for the *P. acutifolius* calli.

light conditions	experiment 1		experiment 2	
	large	small	large	small
16h light/8h dark	24	39	30	44
24h dark	1	7	0	2

Table 4.3: Number of large and small *uidA* expression zones in 30 explants per experiment and per treatment. The calli were co-cultivated with bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) in complete darkness or under 16-h light/8-h dark conditions.

Zambre and colleagues (2002) confirmed these results in more elaborate experiments with callus from two genotypes of *P. acutifolius* and root segments from two ecotypes of *Arabidopsis thaliana*. In these experiments, *uidA* expression correlated highly and positively with the light period used during co-cultivation; it was inhibited severely by darkness and enhanced more under continuous light than under a 16-h light/ 8-h dark photoperiod. Many other *A. tumefaciens*-mediated transformation protocols use specifically dark co-cultivation conditions and it may be useful to reconsider the parameter light in such procedures.

A series of experiments were aimed at a better induction of the *vir* genes. The *vir* genes of the Ti plasmid mediate the transfer of T-DNA from *A. tumefaciens* to plant cells and the steps preceding integration into the plant genome (reviewed in Gelvin, 2000). Transcription of these *vir* genes can be induced by various related phenolic compounds, such as AS (Stachel et al., 1985), and certain sugars act synergistically with these phenolic inducers (Cangelosi et al., 1990). For optimal expression of the *vir* genes an acidic pH is needed (Stachel et al., 1986). Turk and co-workers (1991) found a maximum *vir* induction response at pH 5.3 for an octopine and leucinopine strain and at pH 5.8 for a nopaline and agropine strain. In many systems, the addition of AS in the co-cultivation medium and/or *Agrobacterium* preculture has proven to be beneficial. Also pretreating the explants with AS can enhance the transformation (Guivarc'h et al., 1993; Boase et al., 1998). In our standard transformation protocol, the agrobacteria are precultured for 4 h in the presence of 200 µM AS and co-cultivated in medium with

20 µM AS. To investigate whether AS is needed in the co-cultivation medium to obtain *Agrobacterium*-mediated gene transfer to *P. acutifolius* calli and whether the efficiency can be improved, AS was added at different concentrations (0, 20, 200 and 2000 µM) to the co-cultivation medium. This experiment was repeated more than five times, always giving consistent results. The number of *uidA*-expressing zones increased with increasing AS concentration and reached a maximum at 200 µM AS (fig. 4.7D; table 4.4). When 2000 µM AS was used, less blue staining was observed and the calli had necrotic zones. More than 10 % of the explants did not survive this co-cultivation. A positive effect of AS on *Agrobacterium*-mediated gene transfer to *Phaseolus* cells has been suggested in previous experiments (Becker et al., 1994), where more tumor proliferation on cotyledonary node explants of *P. vulgaris* was observed after co-cultivation with *A. tumefaciens* in the presence of 100 µM AS.

AS concentration	experiment 1		experiment 2	
	large	small	large	small
0 µM	2	30	4	36
20 µM	30	58	34	34
200 µM	46	73	48	60
2000 µM	23	39	20	14

Table 4.4: Number of large and small *uidA* expression zones observed in two representative experiments with 30 calli each. The calli were co-cultivated with bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) in the presence of 0, 20, 200 and 2000 µM AS.

The necessity of an acidic pH during co-cultivation was also tested. In experiments of Turk and co-workers (1991), the nopaline strain showed a maximum response at pH 5.8 and still displayed significant *vir* induction at higher pH. Our co-cultivation medium was prepared with an initial pH of 5.5, which may however change during the co-cultivation period. We determined whether it was necessary to stabilize the pH of the co-cultivation medium during the whole co-cultivation period, which was achieved by adding 20 mM MES. When 20 mM MES was supplemented, the pH of the medium after co-cultivation was still 5.5 to 5.6, but without buffering, the pH rose to 7.2 during co-cultivation. Figure 4.7E and table 4.5 show that this high pH seemed to prevent for the most part gene transfer. Much more *uidA*-expressing zones could be observed on explants co-cultivated in buffered medium.

pH conditions	experiment 1	
	large	small
with buffering	46	73
without buffering	4	16

Table 4.5: Number of large and small *uidA* expression zones on 30 calli co-cultivated with bacteria of the strain C58C1Rif^R(pMP90)(pATARC3-B1b) in co-cultivation medium with and without buffering with 20 mM MES.

Our results are in contrast to the results obtained by Becker and colleagues (1994) in *P. vulgaris* with an octopine-type bacterial strain. Here, buffering with MES had an inhibitory effect, whereas the pH was not a critical factor for *Agrobacterium* infection.

Other factors tested were the age and size of the calli, the density of the calli in co-cultivation medium, and the time between subculturing and co-cultivation, but none of these parameters seemed to affect the efficiency of transformation substantially. Experiments with a higher density of bacterial inoculum (1.6-2.4 OD_{600}) and a longer co-cultivation period (4 days) were also performed. These circumstances prevented proper killing of the bacteria after co-cultivation and were also detrimental for the calli, resulting in a drastically decreased explant survival rate already 4 days after co-cultivation.

Systematic evaluation of various parameters enabled us to improve the transformation procedure and to obtain a very high level of transient *uidA* expression in the calli. This co-cultivation protocol includes the following conditions: (i) an *Agrobacterium* strain with pMP90 as helper plasmid; (ii) harvesting *Agrobacterium* cells in early-log phase; (iii) co-cultivation in a 16-h light/8-h dark photoperiod at 22 °C; (iv) co-cultivation medium buffered at pH 5.5 with 20 mM MES; and (v) co-cultivation in the presence of 200 µM acetosyringone.

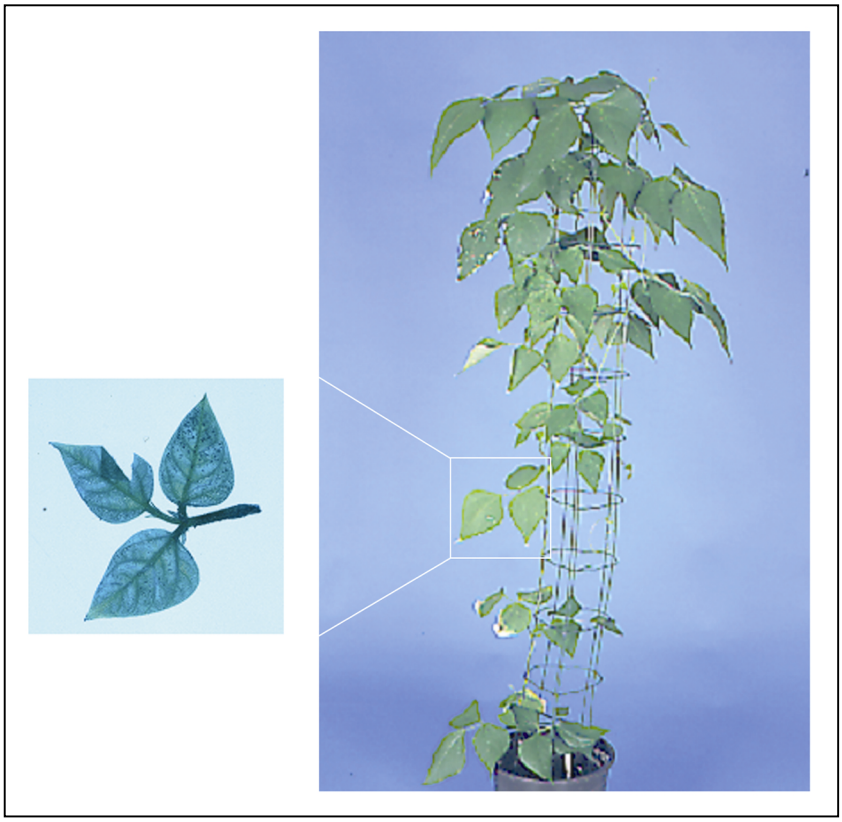


Figure 4.8: GUS activity in a primary transformant of NI576.

Transient expression studies may not always be relevant to stable transformation. In some reports, it is clearly demonstrated that T-DNA transfer from *A. tumefaciens* to the plant cell occurred in an appropriate manner, as demonstrated by efficient transient transformation, but that T-DNA integration seemed to be limiting, resulting in poorly stable transformation. In these cases, improving transient expression levels will not necessarily result in better stable transformation (Shen et al., 1993; Ishida et al., 1996; Maximova et al., 1998). However, most studies show that conditions leading to enhanced transient expression do result in a higher number of transformed plants (e.g., Cao et al., 1998; Kondo et al., 2000; Niu et al., 2000; Zhang et al., 2000; Choi et al., 2001; Trifonova et al., 2001; Suzuki and Nakano, 2002).

Our results also show improved rates of stable transformation with the procedure giving optimum transient expression. Before optimization, one to two independent transformants could be produced in one transformation experiment with 200 explants. When the improved procedure was used, ten independent transgenic lines of the *P. acutifolius* wild genotype NI576 were obtained from a single experiment with 150 explants. After co-cultivation, the callus explants were transferred to selective medium with 20 mg/l geneticin. During four subcultures, each lasting 3 weeks, selection took place. Transgenic shoots developed on these selected callus lines. The number of phenotypically normal shoots produced, varied between the different transgenic lines from 5 to more than 50 shoots. Because survival of rooted shoots was poor, the shoots were *in-vitro* grafted when three trifoliate leaves were produced. From one independent callus line, different clonal transformants could be established. All transgenic plants (fig. 4.8) were fully fertile and produced seeds in quantities comparable with those of non-transformed plants. The time from co-cultivation to obtaining transgenic seeds varied from 9 to 12 months.

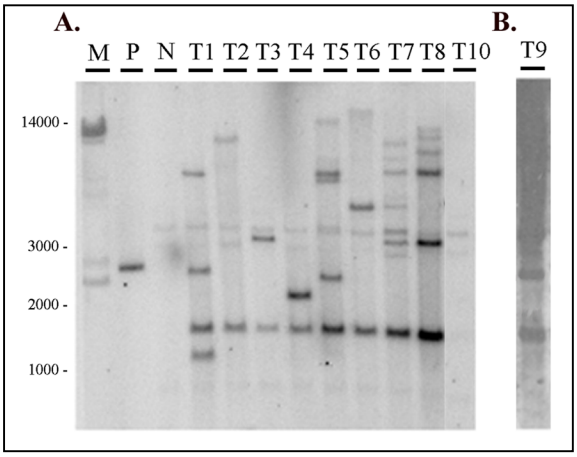


Figure 4.9: Southern analysis of a non-transformed NI576 plant (lane N) and primary transformants (lanes T1 to T10). *Sph*I-digested DNA was probed with an *nptII* fragment hybridizing to an internal 1.4-kb fragment and left-border (LB) fragments in transgenic plants (see fig. 4.11). T2, T3, T4 and T6 probably harbor one T-DNA insertion and T1, T5, T7, T8 and T9 two or more. T10 did not react with the probe. The digest of T9 was not successful in Southern blot A, so part of another blot B is attached. Lane P contains an *nptII* plasmid fragment of 2.5 kb. Lane M contains marker DNA of which the size is indicated in bp on the left.

The stable integration of the introduced transgenes was confirmed by Southern analysis. DNA isolated from leaves of all ten independent primary transformants was digested with *Sph*I and probed with an *nptII* fragment (fig. 4.11). This probe can hybridize to an internal T-DNA fragment of 1.4 kb and to a junction of plant DNA and a left border

fragment of the integrated T-DNA. As shown in figure 4.9, these fragments could be found for 9 out of 10 transformants. Transformants 2, 3, 4, and 6 probably harbor one T-DNA insertion and the other transformants harbor two or more. In plant number 10, the *nptII* gene was not present and also *uidA* expression could not be detected in a histochemical assay (data not shown), suggesting it was not stably transformed. To further confirm transformation, T_1-seeds were harvested and seed proteins were extracted and analyzed by SDS-PAGE and immunoblotting. The binary vector used in this stable transformation experiment, pATARC3-M1, contains between the T-DNA borders the *arc5I-M1* gene coding for the modified Arc5a-M1 protein enhanced with four methionine codons. This protein could be found in seeds of 8 out of 10 transgenic plants (fig. 4.10). As expected, the modified *arc5I-M1* gene was absent in transformant 10. Because the T_1-seeds were not tested for the presence of the *nptII* or *uidA* marker genes before extraction, a negative result could also be related to the segregation of the transgenes, which was probably the case in transformant 2. The relative concentration of Arc5a-M1 in the T_1-seeds varied from 11 to 29 % of the total extracted seed protein. These results confirm the transmission of the transgenes to the offspring.

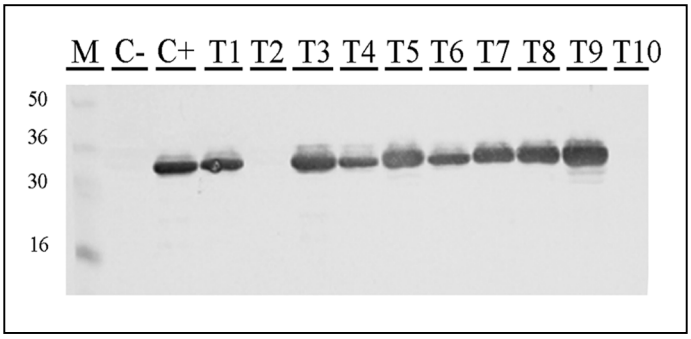


Figure 4.10: Immunoblot analysis of crude protein extracts of seeds of a non-transformed NI576 plant (C-) and NI576 plants transformed with the unmodified *arc5-I* gene (C+) and the modified *arc5I-M1* gene (T1-T10). Lane M contains the marker proteins (molecular mass in kDa).

These results confirm that an improved *Agrobacterium*-mediated transformation system for *P. acutifolius* was established that allows efficient production of transgenic plants. The optimized conditions were also used to produce several series of transformants of the *P. acutifolius* cultivar TB1 (Zambre et al., in preparation). Within the grain legumes, *P. acutifolius* is now one of the few species for which the number of transformed plants that can be generated is large enough to permit transgenic approaches for applied and fundamental research. *P. acutifolius* can be used to test gene engineering strategies, such as improving seed nutritional quality through increasing the methionine content by expressing a modified Arc5a protein (see chapter 5).

The parameters examined in this study only concern the 48 h of the co-cultivation period and should thus have a minimal impact on tissue culture and regeneration response. Similar optimizations may therefore be useful for improving other transformation protocols.

Experimental procedures

Bacterial strains and plasmids

In the initial experiments, the *A. tumefaciens* nopaline-type strain C58C1Rif^R (Holsters et al., 1980) was used that harbors the binary plasmid pTJK136 (Kapila et al., 1997). The following optimization experiments and stable transformation experiment were done with the binary plasmid pATARC3-B1b and pATARC3-M1, respectively. These binary vectors contain all the same P35S-*uidA*-intron-3'*nos* and P*nos*-*nptII*-3'*ocs* cassettes between the T-DNA borders. In addition, pATARC3-B1b contains the unmodified *arc5-I* gene, whereas pATARC3-M1 contains the modified *arc5I-M1* gene (see chapter3, fig. 3.3). Three types of non-oncogenic Ti plasmids were tested: the nopaline-type pMP90 (Koncz and Schell 1986), the octopine-type pGV2260 (Deblaere et al., 1985), and the agropine/succinamopine-type pEHA101 (Hood et al., 1986).

Agrobacterium transformation

All experiments were based on the transformation protocol described above in '*Agrobacterium*-mediated transformation of *P. acutifolius'*, with a few minor modifications. Five days after the third subculture, callus explants were cultured at a density of 15 pieces in 24 ml of co-cultivation medium. Agrobacteria were added to the calli to a final OD_{600} of 0.8, with 1 OD_{600} corresponding to 1.1 ± 0.2 x 10^9 cells/ml for all *Agrobacterium* strains used. After 2 days, explants were washed twice and cultured on non-selective or selective media (20 mg/l geneticin) for optimization and stable transformation experiments, respectively.

Factor evaluation

For each factor tested, at least two independent experiments were performed (with the exception of the pH parameter). In each experiment, a minimum of 30 callus pieces was used per treatment.

To test the effect of the growth phase of the *A. tumefaciens* inocula, bacteria obtained from cultures grown for 36 h (late-log phase, OD_{600} = 2.5 ± 0.3) were compared with those grown for 12 h (early-log phase, OD_{600} = 1.4 ± 0.2). A colony from a freshly streaked plate was used to inoculate 5 ml yeast extract broth medium (YEB) (Grimsley et al., 1986) containing the appropriate antibiotics. The culture was incubated for exactly 12 h at 28°C under continuous shaking (150 rpm). This preculture (OD_{600} = 1.2 ± 0.2) was then diluted 20 times in 100 ml YEB medium in a 500-ml Erlenmeyer and incubated exactly for either 12 h or 36 h under the same conditions.

To examine the influence of light, co-cultivation was compared in darkness and under a light period of 16 h (light intensity 20 µmol/m^2/s). To avoid temperature differences resulting from irradiation heating, the experiment was carried out in a water bath (Dillen et al., 1997b). To obtain dark conditions the jars were wrapped in aluminium foil.

GUS assay

Comparisons of the transient expression levels were made by assaying the expression of the *uidA*-intron gene in the explants 4 days after co-cultivation. GUS activity was localized histochemically as described above, except that incubation lasted only 5 h.

Transformation frequency was evaluated as the total number of blue spots observable with the naked eye. A distinction was made between small spots (< 0.5 mm diameter, see spot marked "s" in figure 4.7B), representing one or a few *uidA*-expressing cells and large spots (> 1 mm diameter, see spots marked "l" in figure 4.7B), representing a complete cell cluster expressing the *uidA* gene.

Stable transformation and regeneration of transgenic plants

All selection and regeneration steps to obtain transgenic *P. acutifolius* plants were performed as described above.

Protein analysis of transgenic plants

For detection of the Arc5a protein, total protein extracts of T_1-seeds were obtained by two successive extractions of ground cotyledon tissue in 10 mM NaCl and 50 mM Gly, pH 2.4 for 30 min at room temperature under continuous shaking. After centrifugation for 10 min at 20 000 g, the pellet was removed and supernatants were pooled. Protein concentrations of the crude extracts were determined by measuring the UV-A_{280} (Goossens et al., 1999a). 20 µg proteins were separated on a 12.5 % polyacrylamide gel using SDS-PAGE and visualized by Coomassie blue staining. Immunoblot analysis of 1 µg of proteins was performed using a rabbit polyclonal antiserum against unmodified arcelin-5 after transferring the proteins to an electroblotting membrane (Problott™, Applied Biosystems, Foster City, California). Accumulation levels were estimated by analyzing a grey-scale image of the immunoblot using the Imagemaster VDS software (Amersham Pharmacia Biotech). The concentration of the Arc5a-M1 proteins was determined relative to the concentration of the positive control B1b-8 (seed of a homozygous transformant accumulating 23.8 % unmodified Arc5a; Goossens et al., 1999a).

DNA analysis of transgenic plants

The presence of the neomycin phosphotransferase II (*nptII*) gene was analyzed by Southern blot analysis. Genomic DNA was extracted from leaves of non-transformed and putative primary transformants using the DNeasy® Plant Mini kit (Qiagen). DNA (7 µg) was digested with *Sph*I (fig. 4.11), fractionated on a 1 % agarose gel by electrophoresis, transferred to a Hybond-N^+ membrane (Amersham Pharmacia Biotech, Little Chalfont, UK) and hybridized at 65 °C with a 0.8-kb *Bam*HI/*Bgl*II *nptII* fragment excised from pGemD (Ingelbrecht et al., 1989). The probe was radiolabeled with ^{32}P using the T7 Quick Ready Prime II kit (Amersham Pharmacia Biotech).

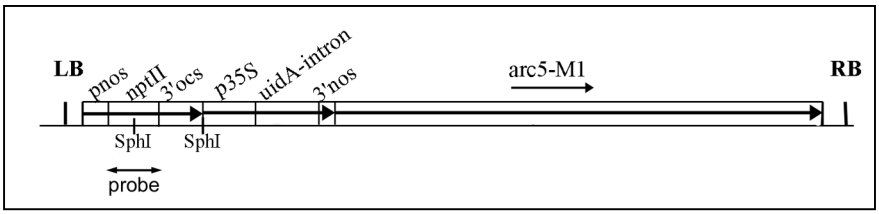


Figure 4.11: The transferred region of plasmid pATARC3-M1. RB, right border; LB, left border*; pnos*, promoter of the nopaline synthase gene; *nptII*, neomycin phosphotransferase II gene; *3'ocs*, 3' signal of octopine synthase, *p35S*, the cauliflower mosaic virus 35S promoter; *uidA*-intron, *Escherichia coli* ß-glucuronidase gene with the potato *st-ls1* intron; *3'nos*, 3' signal of nopaline synthase; *arc5-M1*, gene coding forArc5a-M1. The restriction sites for *SphI* and the *nptII*-probe used for Southern analysis are indicated.

INCREASING THE METHIONINE CONTENT OF PHASEOLUS BEANS

part of the results of this chapter will be submitted together with additional data as 'De Clercq J., Grunewald W., Zambre M., Van Montagu M., Terryn N., Angenon G. Improving the methionine content of *Phaseolus* beans by expressing a modified *P. vulgaris* seed protein arcelin-5a.'

Introduction

Legumes in the genus *Phaseolus*, among which *P. vulgaris* is economically the most important, produce seeds that are consumed worldwide as dietary protein source by millions of people, mainly in Latin America, Africa and India. Like other legume seeds, *Phaseolus* beans contain low amount of the sulfur containing amino acids. The *Phaseolus* seeds harbor only 1.05 g methionine and 0.85 g cysteine per 100 g protein (FAO, 1970), which is less than the FAO reference protein containing 2.5 g per 100 g protein (Scrimshaw et al., 1986). One genetic engineering strategy for changing the amino acid balance of the legume seed storage proteins involves mutagenesis of one of their genes and seed-specific expression of this modified gene in transformed plants (see chapter 1). We have modified the nutritional quality of the *P. vulgaris* seed storage protein arcelin-5a (Arc5a) by the substitution of several amino acids with a methionine and/or insertion of a methionine-rich sequence (see chapter 2). These modified Arc5a proteins can be used to increase the methionine content of *Phaseolus* beans. Accumulation of a modified Arc5a protein with 4 methionines at a level of 20 % of the total seed protein content is theoretically sufficient to elevate the methionine content of *Phaseolus* beans. When a modified Arc5a protein harboring ten additional methionines reaches an accumulation level of 20 % of the total seed protein content, the methionine content will increase to 1.74 g per 100 g protein. As a result, the amount of sulfur-containing amino acids increases from 1.9 g to 2.56 g per 100 g protein and this exceeds the FAO reference value for sulfur containing amino acids.

Stable transformation of most grain legumes, including the *Phaseolus* species, remains very difficult to achieve. Because no *Agrobacterium*-mediated transformation procedure for *P. vulgaris* is available yet, the modified Arc5a proteins were introduced in *P. acutifolius* seeds using the optimized transformation procedure described in chapter 4.

In several grain legumes, over-expression of a high-methionine protein resulted in an increase of protein-bound methionine (soybean, Townsend et al., 1994; narbon bean, Saalbach et al., 1995; narrow leaf lupin, Molvig et al., 1997) without reaching, however, the level which theoretically was expected from the accumulation level of the high-methionine protein in seeds. In transgenic soybean seeds, a strong suppression of several endogenous methionine-containing proteins was observed, indicating that the Brazil nut high-methionine protein was produced at the expense of these endogenous proteins (Muntz et al., 1997). Supplementation of exogenous methionine to cotyledons cultured *in vitro* could reverse these effects (Jung et al., 1997). In narbon bean, on the other hand, the methionine-rich Brazil nut protein was produced at the expense of the secondary metabolite γ-glutamyl-S-ethenyl-cysteine (GEC). GEC was shown to be unpalatable to pigs and this prevents the more widespread use of narbon bean as fodder (Enneking, 1995). Also in transgenic narrow leaf lupin seeds, a decrease in the accumulation of some other sulfur-rich pool(s) is thought to be responsible for the lack of increase in cysteine content and low increase in methionine content. These results suggest that availability of free methionine limits the synthesis of methionine-rich proteins in the seeds.

Methionine is an amino acid derived from aspartate. One of the key enzymes in the aspartate biosynthetic pathway is aspartate kinase (AK), which is feedback inhibited by the amino acids threonine and lysine, end products of this pathway (see chapter 1). Expression of a mutant AK gene from *Escherichia coli*, encoding a feedback-insensitive enzyme, in transgenic tobacco seeds led to an increase in free threonine (16 fold) and methionine (3 fold) (Karchi et al., 1993). Expressing the same mutant AK gene in *Vicia narbonensis* resulted in a 3- to 7-fold increase of free threonine and approximately 2-fold more free methionine (Pickardt et al., 1998).

In order to increase the methionine content of *Phaseolus* beans significantly, we anticipated that a synergistic approach would give the best results. Methionine biosynthesis rates were increased through seed-specific expression of a mutant AK gene, while, at the same time, the expression of a methionine-enriched arcelin-5a protein created a strong methionine sink in the seeds. This was achieved by co-transformation with both genes. Co-transformation was done successfully before, e.g., with *Brassica napus* (De Block et al., 1991) and *A. thaliana* and tobacco (De Buck et al., 1998) where co-transformation frequencies were obtained of 39-85 %, 21-47 % and 41-44 % respectively. Unfortunately, no results of this approach are available since none of the transgenic *P. acutifolius* seeds accumulating the modified Arc5a protein harbored the mutant AK gene.

Results

Introduction of the transgenes into *Phaseolus acutifolius*

Green nodular callus of the *P. acutifolius* wild genotype NI576, established from bud explants (Dillen et al., 1996), was co-cultivated with the *A. tumefaciens* strains C58C1Rif^R (pMP90) (pATARC3-M1) as described in chapter 4 in 'Optimization of the transformation procedure for *P. acutifolius'*. The binary vector pATARC3-M1 contains between the T-DNA borders the two marker genes *nptII* and *uidA*-intron, as well as the modified *arc5I-M1* gene with 4 additional methionine codons. From one transformation experiment with 150 callus pieces, 9 independent transgenic lines were obtained. All transgenic plants produced seeds in quantities comparable with those of non-transformed NI576 plants.

Regeneration-competent callus of the *P. acutifolius* domesticated genotype TB1, established from *in vitro* germinated seeds (Zambre et al., 1998a), was co-cultivated with the *A. tumefaciens* strains C58C1Rif^R (pMP90) harboring the binary vector (pATARC3-M12) or (pATARC3-M12I2). These binary vectors contain the modified *arc5I-M12* gene that codes for an Arc5a protein with 7 additional methionines and the modified *arc5I-M12I2* gene coding for an Arc5a protein with 12 additional methionines, respectively (see chapter 2). For each *A. tumefaciens* strain, one transformation experiment was performed with 30 g of callus resulting in 3 independent transgenic lines for both strains. In addition, a co-transformation experiment was done with the *A. tumefaciens* strains C58C1Rif^R (pMP90) (pATARC3-M12) and C58C1Rif^R (pMP90) (pATAG23-AK). The latter binary vector contains between the left and right T-DNA border the phosphinotricin acetyltransferase (*bar*) marker gene and a mutated aspartate kinase (AK) gene (*lysC*) that codes for a bacterial AK that is insensitive for feedback control by threonine and lysine (Shaul and Galili, 1992). This experiment produced 4 independent transgenic lines resistant to geneticin. All transgenic plants produced seeds, but the quantity and quality of these seeds was not always good due to sub-optimal growth conditions for the transgenic plants in the greenhouse.

Detection of co-transformed plants

The four geneticin-resistant TB1 lines (T1.1-T1.4) that originated from the co-transformation experiment were further tested for resistance to the herbicide phosphinothricin (BASTA®) by spraying leaves *in vivo*. Because all treated leaves died, transgenic T_1-seeds were tested in *vitro* by germinating them on BASTA®-containing medium. Again, no transgenic seeds showed resistance. Seven days after sowing, the transgenic seeds were not germinated yet in contrast with seeds on medium without BASTA® which developed primary leaves and roots at this stage. Also tobacco seeds of a plant transformed with the *bar* gene showed normal germination on the BASTA®-

containing medium. Fourteen days after sowing, still no primary leaves were developed and the cotyledons and small roots coming out of the transgenic seeds showed necrotic zones (fig. 5.1).

Figure 5.1: Evaluation of BASTA® resistance, 14 days after germination. Seedling 1 comes from a wild-type TB1 seed germinated on GM1 without BASTA®, seedling 2 from a wild-type seed germinated on GM1 with 10 mg/l BASTA® and seedling 3 to 6 come from geneticin-resistant transgenic TB1 seeds, co-transformed with the *A. tumefaciens* strains C58C1Rif^R (pMP90) (pATARC3-M12) and C58C1Rif^R (pMP90) (pATAG23-AK), germinated on BASTA®-containing GM1.

Finally, DNA from leaves of these primary transformants was used for PCR analysis with primers amplifying a 510 bp fragment of the *bar* gene (fig. 5.2). This *bar* fragment could be detected in leaves of a BASTA®-resistant tobacco plant, but not in leaves of the *Phaseolus* transformants. All these tests clearly demonstrate that none of the transgenic TB1 lines harbors the *bar* gene indicating that no co-transformation event happened. This implies that there is no transgenic *P. acutifolius* line expressing the mutant *lysC* gene.

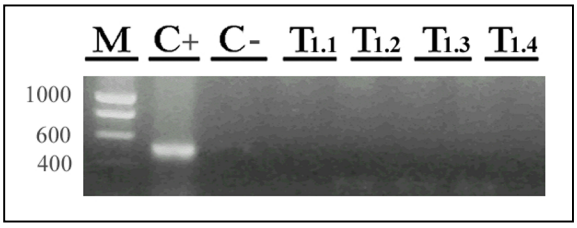


Figure 5.2: PCR analysis of genomic DNA extracted from leaves of a BASTA®-resistant tobacco plant (C+), a wild-type TB1 plant (C-) and primary TB1 geneticin-resistant transformants T1.1-T1.4.

Detection of the modified Arc5a proteins in transgenic seeds

Transgenic T_1-seeds were analyzed for the presence of the modified Arc5a proteins using SDS-PAGE and Immunoblot analysis. For the Immunoblot analysis a rabbit polyclonal antibody against the unmodified Arc5a protein, which also recognizes the modified Arc5a proteins (see chapter 3), was used. The concentration of the modified Arc5a proteins was determined relative to the concentration of unmodified Arc5a found in seeds of the homozygous NI576 line, B1b-8 (23.8 % of the total extractable seed protein; Goossens et al., 1999a). Protein extracts of the latter seeds were used as positive control in SDS-PAGE and Immunoblot analysis.

The modified Arc5a-M1 protein with 4 additional methionines could be detected in the seeds of all nine transgenic *P. acutifolius* NI576 plants (fig. 5.3 and 4.10). The relative concentration of the Arc5a-M1 protein varied from 11 to 29 % of the total extractable seed protein. These results are very similar to the ones obtained with unmodified Arc5a (Goossens et al., 1999a).

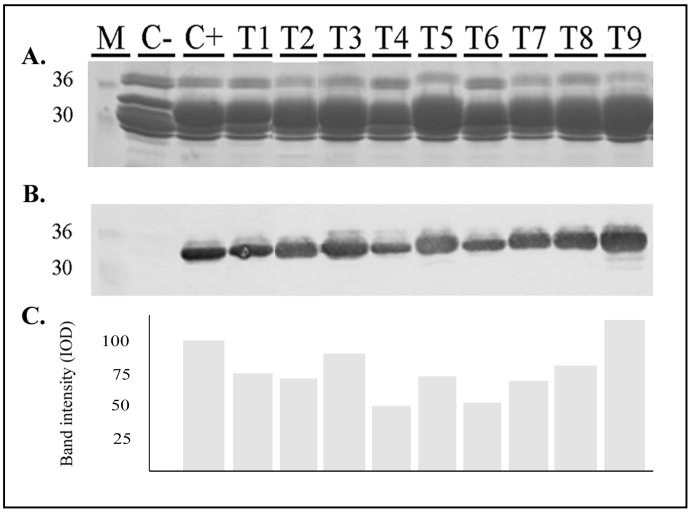


Figure 5.3: SDS-PAGE (A) and Immunoblot (B) analysis of crude seed extracts of non-transformed NI576 plants (C-) and transformed NI576 plants with the unmodified *arc5-I* gene (C+) and the modified *arc5I-M1* gene (T1-T9). Lane M contains the marker proteins of which the molecular mass is indicated in kDa on the left. Quantification of band intensity is shown (C) using the immunoblot with the help of the Imagemaster quantifier.

The modified Arc5a-M12 protein with 7 additional methionines could be detected in transgenic seeds of *P. acutifolius* TB1 plants. Here, the modified Arc5a-M12 accumulated in seeds of six out of seven transgenic plants (fig. 5.4). In transformant 2.1 the modified Arc5a-M12 protein could not be isolated from the seeds, although the transgene was present in leaves of all primary transformants, detected through Southern blot analysis with a probe containing the coding sequence of the *arc5I-M12* gene (data not shown). The relative concentration of the Arc5a-M12 protein in transgenic TB1 seeds varied somewhat more than in the *arc5I-M1* lines, i.e. from 5 to 32 % of the total extracted seed protein. Moreover, there was also considerable variation between different seeds of a single transgenic plant (data not shown). In earlier experiments, Goossens and co-workers (1999a) remarked a low plant-to-plant variation in transgene expression in *P. acutifolius* and *A. thaliana* plants transformed

with the unmodified *arc5* gene. Therefore, we suppose that the fluctuations found in this set of experiments are related to the sub-optimal growth conditions of the transgenic plants at the moment of seed setting.

The modified Arc5a-M12 protein was also detected in progeny (T_2) seeds of one transgenic plant tested (T1.2) at similar accumulation levels (data not shown). This provides clear evidence for transmission of the transgenes to the offspring.

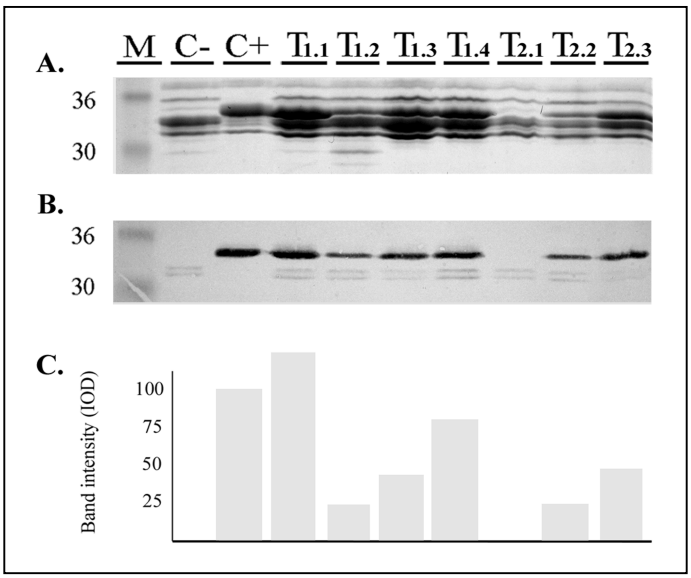


Figure 5.4: SDS-PAGE (A) and Immunoblot (B) analysis of crude seed extracts of a non-transformed TB1 plant (C-), a transformed NI576 plant with the unmodified *arc5-I* gene (C+) and transformed TB1 plants with the modified *arc5I-M12* gene from two different experiments, T1.1-T1.4 from experiment 1 and T2.1-T2.3 from experiment 2. Lane M contains the marker proteins of which the molecular mass is indicated in kDa on the left. Quantification of band intensity is shown (C) using the immunoblot with the help of the Imagemaster quantifier.

The transgenic TB1 plants obtained after transformation with the *arc5I-M12I2* gene produced only a few seeds. These seeds were not analyzed but immediately used to produce offspring. At the moment of writing this thesis, two transgenic lines had produced a few ripe pods of which two seeds were used for analysis. The modified Arc5a-M12I2 protein harboring 7 substitutions with a methionine and a methionine-rich insertion could be detected in these seed extracts at high levels. The banding pattern for this modified protein detected with SDS-PAGE and Immunoblot analysis consists of one band migrating around 33.2 kDa, one or two bands around 18 kDa and one or two bands around 15 kDa. The band of 33.2 kDa represents the full-length protein which became approximately 1 kDa heavier than the unmodified Arc5a protein because of the insertion. The other bands correspond to the banding pattern found in *Arabidopsis thaliana* seeds for this modified protein (see chapter 3, fig. 3.9) and represent some breakdown polypeptides. The relative concentration of the full-length Arc5a-M12I2 protein in the transgenic TB1 seeds varied between 7 and 21 % of the total extractable seed protein. Including the breakdown polypeptides found on the immunoblot, this accumulation level mounts up to 35 % of the total extractable seed protein.

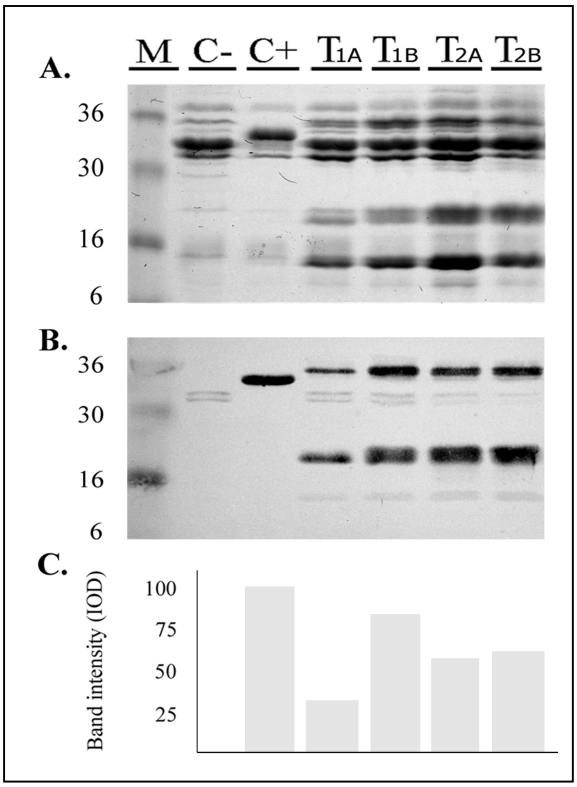


Figure 5.5: SDS-PAGE (A) and Immunoblot (B) analysis of crude seed extracts of a non-transformed TB1 plant (C-), a transformed NI576 plant with the unmodified *arc5-I* gene (C+) and transformed TB1 plants with the modified *arc5I-M12I2* gene. Of each transgenic plant, seed extracts of two different seeds were used (A and B). Lane M contains the marker proteins of which the molecular mass is indicated in kDa on the left. Quantification of band intensity is shown (C) using only the band of the full-length protein on the immunoblot with the help of the Imagemaster quantifier.

Methionine content in transgenic seeds

In a preliminary experiment, the amino acid composition of T_1-seeds of two transgenic TB1 lines harboring the Arc5a-M12 protein and one control TB1 plant was analyzed. As can be seen in table 5.1, the methionine content significantly improved in both transgenic lines. These data should be cautiously interpreted however, because the control and transgenic seeds were grown at a different moment which results in a different composition for most of the amino acids. Further experiments have to be established with seeds grown under the same conditions.

	control	T1.1	T1.4
ASP	15.40 ± 2.07	18.09 ± 1.02	18.61 ± 0.80
GLU	14.19 ± 0.64	19.66 ± 0.78	17.17 ± 1.59
SER	4.11 ± 0.17	4.02 ± 0.11	3.96 ± 0.11
GLY	5.93 ± 0.56	4.41 ± 0.09	4.43 ± 0.09
HIS	1.93 ± 0.01	1.32 ± 0.02	1.53 ± 0.03
ARG	3.22 ± 0.04	2.00 ± 0.07	2.53 ± 0.02
THR	2.10 ± 0.21	1.68 ± 0.07	1.81 ± 0.07
ALA	3.36 ± 0.04	2.80 ± 0.07	3.06 ± 0.08
PRO	4.78 ± 0.60	2.52 ± 0.13	2.82 ± 0.12
TYR	6.79 ± 0.61	5.86 ± 0.10	8.91 ± 0.18
VAL	8.88 ± 0.32	9.99 ± 0.13	9.48 ± 0.17
MET	0.36 ±0.05	0.54 ± 0.13	0.58 ± 0.03
ILE	1.81 ± 0.29	3.77 ± 0.08	2.58 ± 0.04
LEU	7.39 ± 0.12	7.42 ± 0.19	6.88 ± 0.17
PHE	6.24 ± 0.56	6.46 ± 0.20	6.00 ± 0.10
LYS	12.52 ± 0.23	10.48 ± 0.67	9.65 ± 0.39

Table 5.1: Amino acid composition (g/100 g protein) of TB1 seeds of an untransformed plant (control) and two transformed plants with the *arc5I-M12* gene (T1.1/T1.4). Data are the average of four separate hydrolyses ± standard deviation.

Discussion

In order to increase the methionine content of *Phaseolus* beans, we expressed several modified genes coding for a *P. vulgaris* seed storage protein arcelin-5a (Arc5a) with enhanced methionine content (chapter 2) in *P. acutifolius* seeds. The Arc5a-M1 protein with 4 substitutions with a methionine accumulated at very high levels in seeds of nine independent transgenic NI576 lines. The accumulation levels were comparable with these of unmodified Arc5a in transgenic NI576 seeds (Goossens et al., 1999a), indicating that the substitutions didn't affect any of the processes needed for transport through the ER and Golgi apparatus, protein deposition and stable storage in the protein storage vacuoles. In seeds of six transgenic TB1 lines, the Arc5a-M12 protein with 7 substitutions with a methionine accumulated at high levels some of which were comparable to these of the Arc5-M1 and unmodified Arc5a protein. Also the Arc5a-M12I2 protein, with 7 substitutions with a methionine and an insertion with a methionine-rich sequence with 5 additional methionines, was produced at high levels in transgenic TB1 seeds. Although this protein was partially susceptible to degradation, the obtained accumulation levels of the full-length protein, in the four seeds tested, were still comparable with these of unmodified Arc5a and the other modified Arc5a proteins in transgenic *Phaseolus* beans. These data show that achieving an accumulation level of 20 % of the total seed protein content, as needed to obtain a methionine level exceeding the FAO reference value in *P. vulgaris,* is feasible.

In one preliminary experiment, it was shown that the high accumulation levels of Arc5a-M12 resulted in an enhancement of the methionine content of *P. acutifolius* seeds. We indeed expect that expressing this modified gene with 7 additional methionine codons or, even more, the modified *arc5I-M12I2* gene with 12 additional methionine codons can increase the methionine content of *Phaseolus* beans substantially. Theoretically, a 40 % and 80 % increase is expected when the modified Arc5a-M12 and Arc5a-M12I2 protein, respectively, reach an accumulation level of 20 % of the total protein content in *Phaseolus* seeds.

Several publications mention that methionine biosynthesis in seeds could be a limiting factor when increasing the methionine content of seeds through expression of a high-methionine protein. We wanted to anticipate this problem with synergistic expression of the modified *arc5I-M12* gene together with the mutated *lysC* gene. Expression of this feedback-insensitive aspartate kinase gene from *Escherichia coli* in transgenic tobacco seeds (Karchi et al., 1993) and transgenic *Vicia narbonensis* seeds (Pickardt et al., 1998) resulted in a 2- to 3-fold increase of the free methionine content of the seeds. In an attempt to achieve the expression of both genes, co-transformation with two *A. tumefaciens* strains was performed. Unfortunately, none of the four transgenic TB1 plants selected for presence of the *arc5I-M12* gene harbored the *lysC* gene. To test whether the methionine content can be further enhanced by co-expressing both genes, more TB1 plants need to be co-transformed. Another possibility is producing

transgenic TB1 plants expressing only the mutated *lysC* gene in seeds and crossing these plants with transgenic TB1 plants accumulating the modified Arc5a-M12 protein in their seeds.

The final goal is to increase the methionine content of *P. vulgaris* beans, which are used mostly in foods. Because no *Agrobacterium*-mediated transformation procedure is available yet for *P. vulgaris*, *P. acutifolius* was used as a model plant. Furthermore, introduction of transgenes from *P. acutifolius* into *P. vulgaris* can be a way to genetically engineer the latter. Introgression of a simple inherited trait from *P. acutifolius* to *P. vulgaris* is feasible as has been demonstrated for *Xanthomonas campestris* resistance (Scott and Michaels, 1992).

These modified Arc5a proteins with 7 or 12 additional methionines can not only be used to improve the methionine content of *Phaseolus* beans but also of other legumes, which are used for human or animal nutrition.

Experimental procedures

Preparation of the aspartate kinase construct

To construct a binary vector containing the aspartate kinase (*lysC*) gene together with the phosphinotricin acetyltransferase (*bar*) marker gene, several cloning steps were needed (fig. 5.7). First, a new multicloning site was created in the pATAG4 binary vector (Goossens et al., 1999a) using two complementary nucleotide sequences 5' AATTCGTTTAAACGTCGACAGATCTCCTGCAGGAAGCTTCCCGGGGATCCGAGCTCTCTAG ACTCGAGC 3' and 5' GCAAATTTGCAGCTGTCTAGAGGACGTCCTTCGAAGGGCCCCTAGGC TCGAGAGATCTGAGCTCGTCGA 3' (=). This DNA stretch was included after a *HindIII* digest resulting in the loss of the *nptII* marker gene. The new binary vector was called pATAG20 (fig. 5.7A).

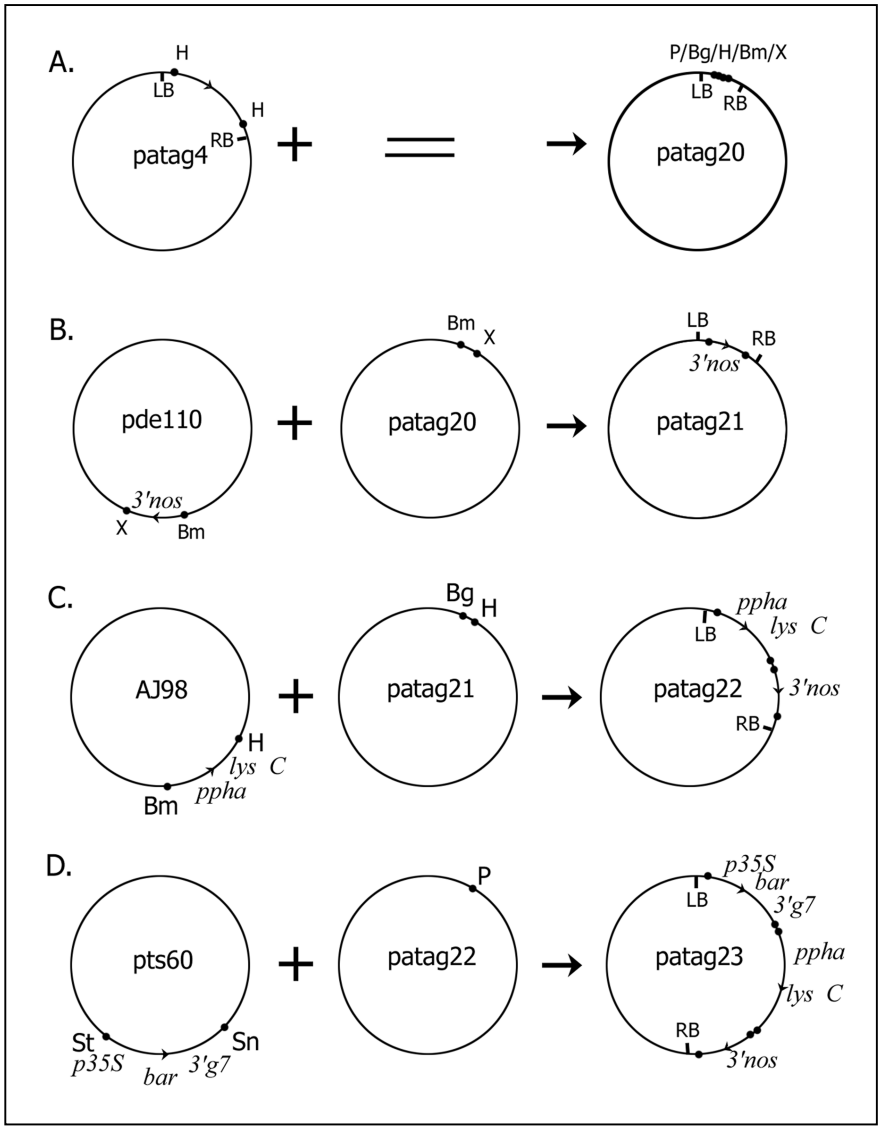


Figure 5.7: Scheme of the different cloning steps needed to construct the binary vector pATAG23-AK (A-D) with indication of the used restriction sites *HindIII* (H), XbaI (X), *BamHI* (Bm), *PmeI* (P), *BglII* (Bg), *SnaBI* (Sn) and *StuI* (St). LB=left border, RB=right border of the T-DNA.

Subsequently, an *XbaI/BamHI* fragment from the pDE110 plasmid (Denecke et al., 1989) containing the 3' termination and adenylation signals of the nopaline synthase gene (3'*nos*) was ligated into *XbaI/BamHI*-cut pATAG20 to produce pATAG21 (fig. 5.7B). The plasmid AJ98 (kindly provided by Gad Galili, Weizman Institute, Rehovot, Israel) harbours a mutated *lysC* gene from *Escherichia coli* coding for a feedback-desensitized aspartate kinase (Shaul and Galili, 1992), driven by the β-phaseolin promoter (*ppha*). This gene with its 5' flanking sequences but without the 3' flanking sequences was excised with *BamHI* and *HindIII* and subcloned into the *BglII* and *HindIII* sites of the pATAG21 in front of the 3'*nos* sequence to give pATAG22 (fig. 5.7C). Finally, the *bar* gene (De Block et al., 1987) under control of the CaMV 35S promoter (*p35S*) and 3' octopine T-DNA gene 7 processing and termination sequences (*3'g7*, Dhaese et al., 1983) was excised out of the pTS60 plasmid (D'Halluin et al., 1992) using *SnaBI* and *StuI* and ligated into the unique *PmeI* site at the required position of the pATAG22 vector creating pATAG23-AK (fig. 5.7D).

Bacterial strains and plasmids

Co-cultivation was carried out with the *A. tumefaciens* strain C58C1Rif^R (pMP90) harboring the binary vector pATARC3-M1, pATARC3-M12 or pATARC3-M12I2 which contains between the left and right T-DNA border the neomycin phosphotransferase II (*nptII*) and β-glucuronidase (*uidA-intron*) marker genes, as well as the modified *arc5I-M1, arc5I-M12* or *arc5I-M12I2* gene respectively (see chapter 3, fig 3.3). In one transformation experiment, co-transformation was performed with the *A. tumefaciens* strain C58C1RifR (pMP90) including the binary vector pATAG23-AK harboring the *bar* marker gene and a mutated *lysC* gene coding for feedback-desensitized aspartate kinase (see above).

Transformation procedure

Green nodular callus from bud explants of NI576 was established and co-cultivated as described in chapter 4.

Green nodular callus of TB1 was established as described by Zambre and co-workers (1998a). Seeds of the *P. acutifolius* cultivar TB1 were sterilized and germinated *in vitro* on GM1 medium (Murashige and Skoog (1962) salts and organic addenda, 30 g/l sucrose, 5 mM 6-benzylaminopurine, 7 g/l agar (Gibco, bacterial agar), pH 5.8). After a 3-day incubation under 16 h light (light intensity 20 µmol/m^2/s) at 22 °C, the seed coat was removed and the cotyledons were gently pulled apart from the embryonic axes. Both were cultured on CIM1/5 (same as CIM1 but with 0.1 mg/l TDZ and 0.05 mg/l IAA) in 100x20 mm Petri Dishes (Sigma) at 25 °C to establish green nodular callus. Plates were incubated in the dark for 1 week and thereafter under 16-h light/ 8-h dark regime. After four passages on CIM1/5, each lasting for approximately three weeks, callus pieces were used for co-cultivation as described by Zambre and

colleagues (in preparation). Three gram of callus, derived from 2-3 well-germinated seeds, was gathered in glass jars (8 cm diameter, 14.5 cm height with transparent with plastic screw cap) in 200 ml co-cultivation medium. The agrobacteria were added to the explants to a final OD_{600} of 0.05. Co-cultivation lasted for 7 days during which the co-cultivated cultures were shaken by hand once a day for a few seconds to avoid formation of bacterial film on the liquid medium. After washing, the calli were cultured on selective medium containing 5 mg/l. In subsequent subcultures, geneticin concentration was always 20 mg/l.

For co-transformation, both *A. tumefaciens* strains C58C1Rif^R (pMP90) (pATARC3-M12) and C58C1Rif^R (pMP90) (pATAG23-AK) were added to the callus explants, each at an optical density (OD_{600}) of 0.025.

All subsequent manipulations to obtain transgenic plants were as described in chapter 4, except that for TB1 experiments geneticin was omitted from SIM and SDS media and carbenicillin was added (50 mg/l) instead of cefotaxime. The explants used for co-transformation were selected on geneticin-containing medium as always. Co-transformation events were traced later on transgenic plants (see BASTA® resistance assay).

Growth conditions of transgenic plants

In the first transformation experiment, the *P. acutifolius* wild-type NI576 was used. Wild-type and transgenic plants of this genotype were grown in a separate growth chamber with short-day conditions (8 h light (50 µmol/m^2/s), 16 h dark) at 25 °C. Subsequent transformation experiments were performed with the *P. acutifolius* cultivar TB1. Transgenic plants of TB1 were grown in a greenhouse without light and minimal temperature regulation, depending on the light and temperature of outside. Because TB1 plants require short-day conditions for flowering, it took more then three months before they started flowering. Moreover, the greenhouse was shared with other plants that were a source of different infections.

Protein analysis of transgenic plants

For detection of the modified Arc5a proteins, SDS-PAGE and Immunoblot analysis was performed as described before, except that seeds were GUS-tested before protein extraction. Accumulation levels were estimated by analyzing a grey-scale image of the immunoblot using the Imagemaster VDS software (Amersham Pharmacia Biotech). The concentration of the Arc5a-M1 proteins was determined relative to the concentration of the positive control B1b-8 (seed of a homozygous transformant accumulating 23.8 % unmodified Arc5a; Goossens et al., 1999a).

DNA analysis of transgenic plants

The presence of the modified *arc5-I* gene was analyzed by Southern blot analysis as described before, except for more stringent washing (0.1 x SSC and 0.1 % SDS). Five to seven µg DNA was digested with *XbaI*, fractionated on gel, transferred to membrane and hybridized at 65 °C with a 1.1-kb *Sac*I/*Xba*I *arc5-I* fragment (fig. 5.8). Labeling and detection was performed with 'the Gene Image Random Prime Labeling Module' and 'the Gene Images CDP-Star Detection Module' system (Amersham Pharmacia Biotech).

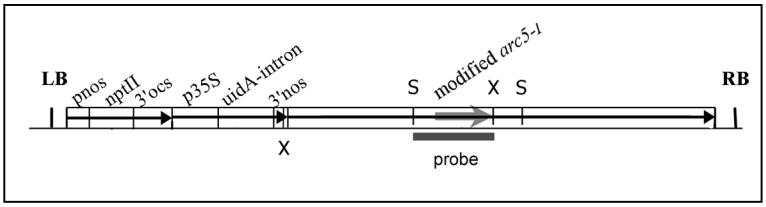


Figure 5.8: The transferred region of the binary vector pATARC3-M12 and pATARC3-M12I2 with indication of the coding sequence for the modified *arc5-I* gene (arrow), the probe used in Southern analysis to detect the presence of the modified *arc5-I* gene (bar) and the restriction sites *SacI* (S) and *XbaI* (X).

To detect the presence of the *bar* gene in the genome of transformed plants, PCR amplification was used. Young leaves were homogenized in 2.0 ml tubes with the help of liquid nitrogen. CTAB buffer (2 % CTAB, 100 mM Tris-HCl pH 7.5, 750 mM NaCl, 10 mM EDTA) was added, followed by incubation at 65 °C for 30 min. The mixture was extracted with 750 µl chloroform/isopropanol (24/1) and DNA was precipitated with 700 µl isopropanol. Approximately 100 ng DNA was used in a 50 µl PCR sample consisting of 5 µl 10 x PCR buffer (Boehringer), 1 unit Taq DNA Polymerase (Boehringer), 2 µl 5 mM dNTPs, and 50 ng sense and antisense primers. The sense and antisense primers used for amplification, yielding a 510 bp internal fragment of the *bar* gene, were *bar rv* 5' CCGCCGTGCCACCGAGGCGGACATGCC 3' and *bar fw* 5' GACGGGCAGGACCGGACGGGGCGG 3'. The cycling program consisted of 35 cycles of amplification comprising a 1 min denaturation step at 94 °C, a 1 min annealing step at 50 °C and a 1 min extension step at 72 °C. A final elongation step of 10 min at 72 °C was included. PCR amplification was performed using the GeneAmp PCR system 2400 (Perkin Elmer).

GUS assay

GUS activity in plant explants was localized histochemically as described above with an incubation period of 5 h.

To test whether a seed harbors the transgenes, a small chip of the cotyledons was tested for GUS activity through incubation for ½ h at 37 °C.

BASTA® resistance assay

Two leaves of each primary transformant were evaluated by spraying a 0.5 % phosphinothricin (PPT = BASTA®, Hoechst Schering AgrEvo A/S, Germany) solution supplemented with 0.1 % Tween-20. The herbicide tolerance was determined one week after treatment. Additionally, BASTA® was evaluated by a germination test. GUS positive seeds of primary transformants were *in vitro* germinated on GM1 (see above) with 10 mg/l BASTA® to look at their ability to grow. Plantlets were observed 7 and 14 days after treatment.

Amino acid analysis

Protein extraction for amino acid analysis was performed as described by Aragão and co-workers (1999). One seed of each transformant was ground in liquid nitrogen to a fine powder. The powder was defatted three times with 25 ml hexane for 10 min at room temperature with periodic agitation, followed by centrifugation at 6000 g for 6 min at 4 °C. The supernatant was discarded and the defatted material was dried in a Speed Vac Concentrator (Savant). After drying, 110 mg powder was dissolved in 880 µl extraction buffer (0.5 M NaCl in 50 mM phosphate buffer, pH 7.2). Extraction was carried out in a 1.5 ml tube with constant rotation at 4 °C for 4 h. The protein extract was clarified through repetitive centrifugation at 9000 tpm at room temperature and the supernatant was dried in the Speed Vac Concentrator. Gas phase hydrolysis was carried out in borosillicate glass tubes placed in a hydrolysis vial, using 6 N HCl as hydrolysis agent for 24 hours at 106 °C. Amino acid composition analysis were performed on a 420A PITC-Derivatizer with on-line 130A Separation System and 920A Data Module (all parts from Applied Biosystems).

SUMMARY AND PERSPECTIVES

Plant seeds represent the primary nutritional protein source for humans as well as animals. The seed protein fraction is, however, unbalanced with reference to the needs of humans as well as domesticated monogastric animals because essential amino acids are present at levels insufficient for dietary requirements. Compared to the FAO reference protein (Scrimhaw et al., 1986), cereal proteins tend to be low in lysine and tryptophan content while legume seed proteins contain low amounts of the sulfur-containing amino acids, methionine and cysteine (FAO, 1970). In cereals, this deficiency could be partly overcome through conventional breeding efforts. The lysine concentration was improved by genetic suppression of the lysine-poor endosperm proteins, in favor of increased amounts of other proteins rich in lysine (Bright and Shewry, 1983). This strategy is not useful in legumes since virtually all known legume seed storage proteins contain low amounts of the sulfur-containing amino acids (Bliss, 1990). Recent developments in recombinant DNA technology offer alternative molecular approaches to modify the amino acid composition of seed proteins.

Several biotechnological strategies have been used to increase the methionine content of legume seeds (see chapter 1). A first approach is engineering the amino acid metabolism of the seed in order to increase the free amount of methionine. This can for example be obtained by expressing genes encoding methionine biosynthetic enzymes that are insensitive to feedback inhibition. Seed-specific expression of a bacterial desensitized aspartate kinase, the first enzyme of the aspartate pathway, increased the production of free methionine two- to three-fold in seeds of *Vicia narbonensis* (Pickardt et al., 1998). A second idea to tackle the problem is influencing the accumulation level of the seed proteins by using the anti-sense approach. This was done successfully in *Brassica napus* (Kohno-Murase et al., 1995). Most probably, this strategy can not be used in legumes because most legume seed storage proteins contain few sulfur-containing amino acids. A third possible strategy is seed-specific expression of heterologous genes coding for high-methionine proteins. Genes for methionine-rich 2S albumins from Brazil nut and sunflower have been used widely for this purpose. Seed-specific expression of the gene from Brazil nut resulted in significant increases of the methionine content in seeds of *Vicia narbonensis* (Pickardt et al., 1995), *Glycine max* (Townsend et al., 1994) and *Phaseolus vulgaris* (Aragão et al., 1999) and the 2S albumin from sunflower improved the seed methionine content of *Lupinus angustifolius* (Molvig et al., 1997). However, alternative methionine-rich proteins will have to be tested since the currently used ones revealed to be allergenic (Nordlee et al., 1996; Kelly and Hefle, 2000). A fourth promising strategy is engineering an endogenous seed storage protein: isolating the gene for a methionine-poor seed protein and modifying its nucleotide sequence so that it encodes a protein with an increased methionine content. Till now, the methionine concentration could not be increased significantly by using this strategy because the expression levels

obtained with these modified genes were generally low and/or the modified protein was susceptible to degradation.

To increase the methionine content of legume seeds in general and *Phaseolus* beans in particular, we chose to modify the *Phaseolus vulgaris* seed storage protein arcelin-5a and to express the modified gene in the seeds. Arcelins are abundant seed storage proteins found in some wild *P. vulgaris* genotypes. They are related to the *Phaseolus* lectin phytohemagglutinin. So far, seven arcelin variants have been identified, of which the arcelin-5 variant is used here. Arcelin-5 consists of two major polypeptides: arcelin-5a (Arc5a) and arcelin-5b (Arc5b) encoded by the *arc5-I* and *arc5-II* genes, respectively (Goossens et al., 1994). An *arc5-I* genomic clone was isolated in our lab (Goossens et al., 1995). The Arc5a protein is a suitable protein for our purpose for several reasons. Firstly, the accumulation level of arcelins is very high, both in the wild beans where they contribute 30 to 40 % of the total protein content, as well as in the cultivated beans in which arcelin was introgressed by breeding. Also in seeds from *P. acutifolius* and *A. thaliana* lines transformed with the *arc5-I* gene, high accumulation levels of the arcelin-5a protein were obtained, up to 24 % and 15 %, respectively (Goossens et al., 1999a). Secondly, the gene encoding arcelin-5a as well as the protein itself are completely characterized (Goossens et al., 1994; 1995; 1999a). Thirdly, the crystallographic structure of arcelin-5a has been completely solved (Hamelryck et al., 1996a), allowing to predict the influence of mutations by computer modeling.

To modify the Arc5a protein, substitutions and insertions with methionine codons were made in the *arc5-I* gene using site-directed mutagenesis (see chapter 2). Arc5a has no methionine residues, but a lot of leguminous lectins and other related proteins do have one or more methionine residues. By aligning the amino acid sequences of these homologous proteins, candidate amino acids for replacement by methionine could be identified. Therefore, all positions in the Arc5a amino acid sequence where a leucine, isoleucine, valine or phenylalanine was found, were analyzed. When at these positions one or more methionines and/or mostly hydrophobic amino acids were present in the homologous proteins, the amino acid at that position in Arc5a was taken as a candidate for substitution. In a next step, the location of these sites in the crystallographic structure of Arc5a was explored. Only candidates with a good accessibility situated near the surface of the protein were retained. Finally, after evaluating the influence of the mutations on the protein stability by computer simulations, fifteen positions in Arc5a were retained for methionine substitution. Subsequently, four different modified *arc5-I* genes were designed, each with 3 or 4 methionine substitutions. In addition, two constructs with a methionine-rich insertion in a variable loop at the surface of the protein were created. This variable loop was replaced with a loop consisting of 15 amino acids, of which five are a methionine.

The stability of the modified Arc5a proteins was tested experimentally by transformation of *Arabidopsis thaliana* (see chapter 3). *A. thaliana* was considered an

appropriate host plant because the unmodified *arc5-I* gene is expressed at high levels and in a seed-specific way in this plant (Goossens et al., 1999a).

The modified Arc5a-M1 protein with a methionine residue at the position of Leu17, Leu41, Val44 and Leu47, and the Arc5a-M2 protein with replacement of the amino acids Ile58, Phe73 and Ile75, accumulated at high levels in the transgenic *A. thaliana* seeds and showed similar characteristics as the unmodified Arc5a. Although every mutation was checked by computer simulations, the other groups of substitutions destabilized the Arc5a protein. The Arc5a-M4 protein with substitutions at Phe148, Leu190, Val194 and Leu219, accumulated only at very low levels in *A. thaliana* seeds and the modified Arc5a-M3 protein (substitutions at Leu91, Val94, Phe104 and Ile138) could not be detected in the transgenic *A. thaliana* seeds although the transgene was detected through Southern analysis. Destabilization of Arc5a-M3 and Arc5a-M4 is most probably caused by only one or two of the introduced mutations.

The two other modified Arc5a proteins, Arc5a-MI1 harboring a methionine-rich insertion in the loop Gly35-Arg43 and Arc5a-MI2 with replacement of the loop Asn111-Ans116, accumulated at high levels in *A. thaliana* seeds but both underwent one or more cleavages or other modifications during biosynthesis, maturation or storage in the protein storage vacuoles. For our goal, increasing the methionine content of the seeds, this is not a problem. As long as all cleavage or degradation products are present in the plants seeds, they will contribute to the raising of the methionine content.

Hereafter, the groups of mutations generating highly accumulating methionine-enhanced proteins were combined in five additional modified Arc5a proteins in order to increase the methionine content of the Arc5a protein as much as possible. Of these modified Arc5a proteins, only the Arc5a-M12, harboring seven substitutions and Arc5a-M12I2 with 12 additional methionines, could be detected in transgenic *A. thaliana* seeds. The other modified proteins harboring the substitutions of group M2 and/or M1 together with the insertion situated in the loop Gly35-Arg43 (Arc5a-M12I1, Arc5a-M2I12 and Arc5a-M12I12), could not be detected in the transgenic seeds. The modified Arc5-M12 protein accumulated at high levels and showed similar characteristics as the unmodified Arc5a and the modified Arc5a-M1 and Arc5a-M2 proteins. Also Arc5a-M12I2 accumulated in *A. thaliana* seeds but underwent several cleavages or modifications similar to the Arc5a-I2 protein.

The final goal of this project is to contribute to the development of a strategy to improve the methionine content of *Phaseolus* beans. Within the genus *Phaseolus*, there are five cultivated species (*P. vulgaris*, *P. acutifolius*, *P. polyanthus*, *P. coccineus*, and *P. lunatus*) of which *P. vulgaris* is economically the most important. *Phaseolus* beans are an important food source providing energy and proteins for millions of people, mostly in Latin America, Africa and India. The world production of *Phaseolus* beans goes beyond 20 million ton per year (Parker, 1995; Singh, 1999). Like other legume seeds, *Phaseolus* beans contain low amounts of the sulfur-containing amino acids. They

harbor only 1.05 g methionine and 0.85 g cysteine per 100 g protein (FAO, 1970), which is less than the FAO reference protein containing 2.5 g per 100 g protein (Scrimshaw et al., 1986). Stable transformation of most grain legumes, including the *Phaseolus* species, remains difficult to achieve, however. For *P. vulgaris*, so far, transgenic plants have been obtained only through particle bombardment of seedling apical meristems (Russell et al., 1993; Aragão et al., 1996; 1999). The transformation frequencies obtained with this method are relatively low. For *P. acutifolius,* an *Agrobacterium tumefaciens*-mediated transformation procedure was developed in our lab for the wild genotype NI576 (Dillen et al., 1997a). To improve this transformation procedure, the effect of different factors on T-DNA transfer was examined by measuring transient expression levels of an intron-containing *uidA* gene (see chapter 4). Improved transformation frequencies were obtained by using an *A. tumefaciens* strain carrying nopaline-type virulence genes and with *Agrobacterium* cells in the early-log growth phase. Optimized co-cultivation was performed at 22 °C under a 16-h photoperiod in an acidic medium (pH 5.5) in the presence of 200 µM acetosyringone. By combining the best treatments, an efficient and reproducible transformation procedure was established. Using the improved procedure, ten independent transgenic lines of the *P. acutifolius* wild genotype NI576 were obtained from a single experiment with 150 explants. Southern and Immunoblot analysis confirmed the stable integration and expression of the transgenes in the primary transgenic plants and their progeny. The optimized conditions were also used to produce several series of transformants of the *P. acutifolius* cultivar TB1 (Zambre et al., in preparation).

In order to increase the methionine content of *Phaseolus* beans, three modified Arc5a proteins, highly accumulating in *A. thaliana* seeds, were introduced into *P. acutifolius* using the optimized transformation procedure (see chapter 5). The Arc5a-M1 protein with 4 substitutions with a methionine, accumulated at very high levels in seeds of nine independent transgenic NI576 lines. The accumulation levels were comparable with these of unmodified Arc5a in transgenic *P. acutifolius* seeds (Goossens et al., 1999a), indicating that the substitutions didn't affect any of the processes needed for transport through the ER and Golgi apparatus, protein deposition and stable storage in the protein storage vacuoles. In seeds of six transgenic TB1 lines, the Arc5a-M12 protein with 7 substitutions with a methionine also accumulated at high levels, some of which were comparable to the levels of the Arc5-M1 and unmodified Arc5a protein. Also the Arc5a-M12I2 protein, with 7 methionine substitutions and an insertion with a methionine-rich sequence with 5 additional methionines, was produced at high levels in transgenic TB1 seeds. Although this protein was partially susceptible to degradation, the obtained accumulation levels of the full-length protein were still comparable with these of unmodified Arc5a and the other modified Arc5a proteins in transgenic *Phaseolus* beans.

In one preliminary experiment, it was shown that the high accumulation levels of Arc5a-M12 resulted in an enhancement of the methionine content of *P. acutifolius* seeds. We indeed expect that expressing this modified gene with 7 additional methionine codons or, even more, the modified *arc5I-M12I2* gene with 12 additional

methionine codons can increase the methionine content of *Phaseolus* beans substantially. Theoretically, a 40 % and 80% increase is expected when the modified Arc5a-M12 and Arc5a-M12I2 protein, respectively, reach an accumulation level of 20 % of the total protein content in *Phaseolus* seeds.

If the Arc5a-M12I2 protein does not accumulate at high levels or if some of the methionine residues are lost as a result of cleavage or partial degradation of the protein, the least risk-bearing substitutions from group M3 and group M4 (like Val94, Ile138 and Leu219) can still be used in combination with the substitutions of group M1 and M2 to create a new modified Arc5a protein. In this way, it must be possible to create a modified Arc5a protein with ten additional methionines. Accumulation of such a protein in *P. vulgaris* beans at a level of 20 % of the total seed protein content, should result in a methionine content exceeding the FAO reference value for sulfur containing amino acids, assuming that the cysteine content will not be changed.

A synergistic approach can be envisaged to change not only the demand for methionine but also the supply of methionine in the seeds, which might be a limiting factor for biosynthesis of a methionine-rich protein. As described in chapter 1, the biosynthesis of methionine is a complicated pathway with many important key enzymes and components. To enhance the supply of methionine, some of the enzymatic steps at the beginning or at branch points of the methionine pathway can be influenced (see fig. 1.5).

The first enzyme that can be modified is the aspartate kinase enzyme, which is feedback inhibited by the amino acids threonine and lysine, end products of this pathway. Expression of a mutant feedback-insensitive AK gene from *Escherichia coli* in transgenic tobacco seeds led to an increase in free threonine (16 fold) and methionine (3 fold) (Karchi et al., 1993). Expressing this mutant AK gene in *Vicia narbonensis* resulted in a 3- to 7-fold increase of free threonine and approximately 2-fold more free methionine (Pickardt et al., 1998). We intended to express the same mutant AK gene in *P. acutifolius* together with a modified Arc5a protein via transformation with two *Agrobacterium* strains. Co-transformants were not obtained in this experiment, however (see chapter 5).

Other enzymes may be considered as well. Cystathionine *γ*-synthase (CGS), the first enzyme of *de novo* methionine biosynthesis, seems to play an important role in methionine biosynthesis. In *Arabidopsis thaliana*, excess amounts of methionine reduce the abundance of CGS mRNA by destabilizing it and hence reduce CGS activity (Chiba et al., 1999). An *Arabidopsis* mutant (*mto-1*) was isolated, harboring a mutation that enhances the stability of the CGS mRNA (Inaba et al., 1994). This mutant accumulates up to 40-fold more free methionine in leaves and 5- to 8- fold more in immature fruits. Also in *Arabidopsis thaliana*, over-expression of the CGS gene in transgenic *Arabidopsis* plants led to high accumulation of soluble methionine (up to 20-fold) in most parts of the plant, including seeds (Kim et al., 2002). Because CGS and threonine synthase (TS) compete for their common substrate *O*-

phosphohomoserine (OPH), similar results can be obtained when lowering the activity of TS. Reducing the endogenous TS activity by antisense inhibition in potato resulted in decreased leaf threonine contents while methionine levels were increased tremendously by 60- to 200-fold (Zeh et al., 2001). Tubers contained a methionine level increased by a factor up to 30 and no reduction in threonine.

In addition to increasing the C-flux towards methionine, sulfur assimilation and cysteine biosynthesis can be enhanced to stimulate methionine biosynthesis since cysteine is needed as component in the first step of *de novo* methionine synthesis. The enzyme serine acetyltransferase (SAT) is recognized as prime regulation point in cysteine biosynthesis (Saito, 1999). Over-expression of a bacterial SAT gene in transgenic tobacco and potato led to 2- to 3-fold increase of the leaf cysteine content (Blaszczyk et al., 1999; Harms et al., 2000). Another important aspect is the availability of sulfur. Tabe and Droux (2002) experienced that uptake of sulfate into the cotyledon was the ultimate limit to sulfur accumulation in transgenic lupin seeds transformed with the methionine-rich sunflower albumin. A member of a family of sulfate transporters has recently been implicated in the process of S-delivery to seeds in *Arabidopsis* (Awazuhara et al., 2001). Increasing the activity of such sulfate transporters might enhance the methionine content in seeds as well.

Methionine serves also as precursor for S-adenosylmethionine (SAM) and S-methylmethionine (SMM). In order to increase the amount of free methionine in the seeds, reducing these conversions can be considered. Co-suppression of SAM synthetase, the enzyme responsible for the production of SAM, in tobacco resulted in a 440-fold increase of free methionine in leaves, but this was accompanied by morphological changes and a disagreeable odor (Boerjan et al., 1994).

In the end, it seems that plant metabolic pathways are much more sophisticated than initially thought. Interactions between different organs of the plant and different compartments within a plant cell lead to an enormous complexity. Different pathways influence each other by stimulating or inhibiting certain intermediary steps. Changing one element out of this complex entity can result in multiple changes with severe consequences. Before changing the biosynthetic pathway of methionine, it would be wise to try to understand the entire regulation pathway first.

NEDERLANDSE SAMENVATTING

Zaden van planten vormen een belangrijke eiwitbron in zowel de menselijke als dierlijke voeding. Wanneer deze zaden als enige eiwitbron worden gebruikt, veroorzaakt dit echter problemen omdat de essentiële aminozuren aanwezig zijn in ontoereikende hoeveelheden om in de dagelijkse behoeften te voorzien. De zaadeiwitten van graangewassen vertonen een tekort aan lysine en tryptofaan en de zaadeiwitten van vlinderbloemigen hebben een tekort aan de zwavelhoudende aminozuren, methionine en cysteïne (FAO, 1970), ten opzichte van het FAO referentieproteïne (Scrimhaw et al., 1986). Bij graangewassen kon de hoeveelheid van het limiterende aminozuur lysine verhoogd worden door selectie voor mutanten die de zaadeiwitfracties arm aan lysine verminderd aanmaken ten voordele van de andere eiwitfracties die wel aanzienlijke hoeveelheden lysine bezitten (Bright en Shewry, 1983). Een gelijkaardige strategie lijkt bij vlinderbloemigen niet mogelijk gezien ongeveer alle gekende zaadeiwitten van vlinderbloemigen weinig methionine en cysteïne bevatten (Bliss, 1990). De ontwikkeling van recombinant DNA technieken maken echter alternatieve wegen voor wijziging van de aminozuursamenstelling mogelijk.

Er werden reeds meerdere strategieën toegepast om het methioninegehalte van zaden te verhogen via biotechnologie (zie hoofdstuk 1). Een eerste strategie is verhoging van het vrije methioninegehalte in de zaden door wijziging van de aminozuurbiosynthese. Dit kan bijvoorbeeld door het tot expressie brengen van een gen coderend voor een enzyme uit de methionine biosynthese dat niet meer onderhevig is aan feedback inhibitie. Zo leidde zaadspecifieke expressie van een mutant aspartaatkinase gen uit *Escherichia coli* tot een vrij methionine gehalte in *Vicia narbonensis* zaden dat 2 tot 3 keer hoger lag (Pickardt et al., 1998). Een tweede benadering is het accumulatieniveau van een eigen zaadeiwit wijzigen door gebruik te maken van de antisense strategie. Hierdoor kan de zaadeiwitfractie arm aan S onderdrukt worden zodat een andere zaadeiwitfractie, rijker aan S, hoger geaccumuleerd kan worden. Deze methode werd succesvol toegepast bij *Brassica napus* (Kohno-Murase et al., 1995) maar kan wellicht niet gebruikt worden bij vlinderbloemigen aangezien praktisch alle vlinderbloemige zaadeiwitten zeer weinig zwavelhoudende aminozuren bevatten. Een derde manier is transfer en zaadspecifieke expressie van een heteroloog gen dat codeert voor een zwavelrijk eiwit. Het 2S albumine uit *Bertholletia excelsa* dat 18 % methionine bevat, en dit van *Helianthus annuus* (16 % methionine) werden hiervoor veelvuldig gebruikt. Zaadspecifieke expressie van het gen coderend voor het zwavelrijke 2S albumine uit *Bertholletia excelsa* leidde tot een significante methionine verhoging in zaden van *Vicia narbonensis* (Pickardt et al., 1995), *Glycine max* (Townsend et al., 1994) en *Phaseolus*

vulgaris (Aragão et al., 1999) en dat van *Helianthus annuus* zorgde voor een verbetering van het methioninegehalte in zaden van *Lupinus angustifolius* (Molvig et al., 1997). Onderzoek wees echter uit dat deze eiwitten allergische reacties veroorzaken waardoor andere methioninerijke eiwitten zullen moeten getest worden in de toekomst (Nordlee et al., 1996; Kelly en Hefle, 2000). Een laatste mogelijke strategie is modificatie van een eigen zaadeiwit: isolatie van een gen coderend voor een methionine-arm zaadeiwit en verandering van de nucleotide sequentie zodat het gaat coderen voor een proteïne met een verhoogd methioninegehalte. Tot nu toe waren de expressieniveaus behaald met deze strategie echter laag en/of was het gewijzigde eiwit gevoelig voor afbraak waardoor het methioninegehalte niet noemenswaardig steeg.

Met als doel het methionine gehalte van vlinderbloemige zaden in het algemeen en *Phaseolus* bonen in het bijzonder te verhogen werd door ons geopteerd voor modificatie en productie van het *Phaseolus vulgaris* zaadeiwit arceline-5a in de zaden. Arcelines zijn abundante zaadproteïnen die gevonden worden in sommige wilde genotypen van *P. vulgaris*. Ze zijn verwant met phytohemagglutinin, de *Phaseolus* lectine. Tot nu toe werden zeven arceline varianten beschreven, waarvan de arceline-5 variant hier gebruikt wordt. Het arceline-5 proteïne bestaat uit twee belangrijke polypeptiden: arceline-5a (Arc5a) gecodeerd door het *arc5-I* gen en arceline-5b (Arc5b) gecodeerd door het *arc5-II* gen (Goossens et al., 1994). Een genomische klone die het *arc5-I* gen bevat, kon geïsoleerd worden in ons labo (Goossens et al., 1995). Er zijn meerdere redenen waarom in dit project voor modificatie van het Arc5a proteïne gekozen werd. Eerst en vooral omdat arcelines zeer abundante eiwitten zijn, zowel in de wilde bonen waar ze 30-40 % van het totale eiwitgehalte uitmaken, als in de cultivars waarin de arcelines werden ingekruist. Ook in zaden van *P. acutifolius* en *Arabidopsis thaliana* planten getransformeerd met het *arc5-I* gen werden hoge expressieniveaus behaald, respectievelijk tot 24 % en tot 15 % van het totale zaadeiwitgehalte (Goossens et al., 1999a). Een tweede reden is de complete karakterisatie van zowel het gen dat voor Arc5a codeert als het eiwit zelf (Goossens et al., 1994; 1995; 1999a). Een andere belangrijke reden is dat de driedimensionale structuur van Arc5a reeds gekend is (Hamelryck et al., 1996a) wat toelaat de invloed van de mutaties op de structuur van het Arc5a eiwit na te gaan door computersimulaties.

Om het methioninegehalte van het Arc5a eiwit te wijzigen werden er substituties en inserties gemaakt met methionine codons in het *arc5-I* gen met behulp van plaatsspecifieke mutagenese (zie hoofdstuk 2). Arc5a heeft geen methionines, maar veel andere vlinderbloemige lectines en andere verwante eiwitten bezitten wel één of meerdere methionines. Door aliniëring van de aminozuursequenties van deze homologe eiwitten konden kandidaat aminozuren voor vervanging geïdentificeerd worden. Hiervoor werden alle posities uit de aminozuursequentie van Arc5a gehaald

waar een leucine, isoleucine, valine of phenylalanine voorkomt en één voor één onderzocht. Wanneer in de aliniëring één of meerdere methionines en/of hoofdzakelijk hydrofobe aminozuren voorkwamen op deze positie in de homologe eiwitten, werd het aminozuur op die positie in het Arc5a eiwit als kandidaat voor vervanging beschouwd. In een volgende stap werd de situering van deze kandidaat aminozuren in de kristallografische structuur van Arc5a bekeken. Enkel kandidaten met een goede bereikbaarheid dicht bij het oppervlak van het eiwit werden weerhouden. Nadat de invloed van de vervanging van deze kandidaat aminozuren met een methionine op de stabiliteit van het eiwit was nagegaan met computersimulaties, bleven nog 15 kandidaten over voor vervanging met een methionine. Uiteindelijk werden er vier verschillende gemodificeerde *arc5-I* genen samengesteld met elk 3 of 4 methionine vervangingen. Hiernaast werden ook twee gewijzigde *arc5-I* genen gemaakt met een methioninerijke insertie in een variabele loop aan het oppervlak van het eiwit. Deze variabele loop werd vervangen door een sequentie van 15 aminozuren, waarvan vijf methionines.

De stabiliteit van de gewijzigde Arc5a eiwitten werd experimenteel getest door transformatie van *Arabidopsis thaliana* (zie hoofdstuk 3). *A. thaliana* werd beschouwd als een geschikte gastheer omdat het ongemodificeerde *arc5-I* gen hoog en zaad-specifiek tot expressie gebracht kon worden in deze plant (Goossens et al., 1999a).

Het gewijzigde Arc5a-M1 proteïne met een methionine op de plaats van Leu17, Leu41, Val44 en Leu47, en het Arc5a-M2 eiwit met vervanging van de aminozuren Ile58, Phe73 en Ile75, accumuleerden hoog in transgene *A. thaliana* zaden en vertoonden gelijke eigenschappen als het ongewijzigde Arc5a. Niettegenstaande alle mutaties op voorhand gecontroleerd werden met computersimulaties, zorgden de substituties uit de andere groepen voor destabilisatie van het Arc5a eiwit. Het Arc5a-M4 eiwit met vervanging van Phe148, Leu190, Val194 en Leu219, accumuleerde slecht zeer laag in *A. thaliana* zaden en het gewijzigde Arc5a-M3 proteïne (substituties van Leu91, Val94, Phe104 en Ile138) kon niet teruggevonden worden in de transgene *A. thaliana* zaden hoewel het transgen werd gedetecteerd door Southern analyse. De destabilisatie van deze Arc5a-M3 en Arc5a-M4 eiwitten wordt wellicht slechts door een of twee van de binnengebrachte mutaties veroorzaakt.

De twee andere gemodificeerde Arc5a eiwitten, Arc5a-MI1 met een methioninerijke insertie in de loop Gly35-Arg43 en Arc5a-MI2 met vervanging van de loop Asn111-Ans116, accumuleerden hoog in *A. thaliana* zaden maar beiden ondergingen één of meerdere splitsingen of andere modificaties tijdens biosynthese, transport door het ER of Golgi apparaat of opslag in de eiwitopslagvacuoles. Voor ons doeleinde, verhoging van het methioninegehalte van zaden, is dit geen probleem. Zolang alle splitsings- en/of afbraakproducten aanwezig blijven in de zaden zullen deze bijdragen tot de verhoging van het methionine gehalte.

In een volgende stap werden alle groepen mutaties die aanleiding gaven tot een hoog-accumulerend eiwit met een verhoogd methionine gehalte gecombineerd in vijf nieuwe gemodificeerde Arc5a eiwitten met de bedoeling het methioninegehalte van het gewijzigd Arc5a eiwit zo goed mogelijk te verbeteren. Van deze gewijzigde Arc5a eiwitten kon enkel het Arc5a-M12 eiwit met 7 substituties en Arc5a-M12I2 met 12 methionines teruggevonden worden in de trangene *A. thaliana* zaden. Het gewijzigde Arc5-M12 eiwit werd in grote hoeveelheden teruggevonden in de zaden en het vertoonde gelijkaardige eigenschappen als het ongemodificeerde Arc5a en het gewijzigde Arc5a-M1 en Arc5a-M2 eiwit. Arc5a-M12I2 accumuleerde ook hoog maar dit eiwit onderging meerdere splitsingen of modificaties, net als het Arc5a-I2 proteïne. De andere eiwitten die de substituties van groep M2 en/of M1 samen met de insertie in de eerste loop Gly35-Arg43 bevatten (Arc5a-M12I1, Arc5a-M2I12 and Arc5a-M12I12), konden onder de gebruikte condities niet teruggevonden worden in de transgene zaden.

Het uiteindelijke doel van dit project is bijdragen tot het ontwikkelen van een strategie om het methionine gehalte van *Phaseolus* bonen te verhogen. Binnen het genus *Phaseolus* zijn er vijf gecultiveerde species nl. *P. vulgaris, P. acutifolius, P. polyanthus, P. coccineus* en *P. lunatus* waarvan *P. vulgaris* economisch gezien de meest belangrijke species is. *Phaseolus* bonen behoren tot de belangrijkste grootzadige vlinderbloemigen met een wereldproductie van meer dan 20 miljoen ton per jaar (Parker, 1995; Singh, 1999). Ze vormen de belangrijkste eiwitbron in de voeding van de bevolking van Afrika, Latijns Amerika en India. Net als de andere vlinderbloemige zaden bezitten *Phaseolus* bonen zeer weinig zwavelhoudende aminozuren. Ze bevatten 1,05 g methionine en 0,85 g cysteïne per 100 g eiwit (FAO, 1970) terwijl de FAO referentiewaarde voor zwavelhoudende aminozuren 2,5 g per 100 g eiwit bedraagt (Scrimshaw et al., 1996). Klassieke veredeling methodes slaagden er tot nu toe niet in om het methionine gehalte van deze zaden te verhogen. Stabiele transformatie van de meeste grootzadige vlinderbloemigen, *Phaseolus* species inclusief, blijft echter zeer moeilijk. Tot nu toe werd stabiele transformatie bij *P. vulgaris* enkel bekomen door beschieting met microprojectielen van apikale meristemen van zaailingen (Russell et al., 1993; Aragao et al., 1996, 1999). De transformatiefrequenties waren hierbij echter zeer laag. Voor *P. acutifolius* werd er wel een transformatiemethode ontwikkeld via *Agrobacterium tumefaciens*-infectie, nl. voor het wild genotype NI576 (Dillen et al., 1997a). Om dit transformatieprotocol te optimaliseren, werd de invloed van meerdere parameters nagegaan op de efficiëntie van T-DNA transfer met behulp van het *uidA-intron* merkergen systeem (zie hoofdstuk 4). Verhoogde transformatiefrequenties werden bekomen door gebruik te maken van een *A. tumefaciens* stam met virulentiegenen van het nopaline type en wanneer deze *Agrobacterium* cellen zich in de vroege log groeifase bevonden. Optimale co-cultivatie gebeurde in zuur medium (pH 5.5) in de aanwezigheid van 200 µM acetosyringone, in een groeikamer waar het 16 h licht was en de temperatuur 22 °C bedroeg. Door alle optimale condities te combineren, werd een efficiënte en reproduceerbare

transformatieprocedure bekomen. Met behulp van deze procedure konden in één transformatie-experiment tien onafhankelijke transgene lijnen bekomen worden voor *het P. acutifolius* wild genotype NI576, vertrekkende van 150 explantaten. Stabiele integratie en overerving van de transgenen werd bevestigd met Southern en Immunoblot analyses. Deze geoptimaliseerde condities werden eveneens gebruikt in meerdere transformatie-experimenten om tientallen transformanten te bekomen van de *P. acutifolius* cultivar TB1 (Zambre et al., in voorbereiding).

Om het methioninegehalte van *Phaseolus* bonen te verhogen werden drie gewijzigde Arc5a eiwitten, die hoog accumuleerden in *A. thaliana* zaden, binnengebracht *in P. acutifolius* (zie hoofdstuk 5). Het Arc5a-M1 eiwit met 4 substituties met een methionine, werd in grote hoeveelheden teruggevonden in zaden van negen onafhankelijke transgene NI576 lijnen. Het accumulatieniveau was vergelijkbaar met dat van ongewijzigd Arc5a in transgene *P. acutifolius* zaden (Goossens et al., 1999a), wat erop wijst dat de substituties de biosynthese, het transport doorheen het ER en Golgi apparaat, de proteïne depositie en een stabiele opslag in de eiwitopslagvacuoles niet ongunstig beïnvloedde. Het Arc5a-M12 eiwit met 7 methionine vervangingen werd in zaden van zes transgene TB1 lijnen teruggevonden in grote hoeveelheden. Sommige accumulatieniveaus van dit eiwit waren zeker vergelijkbaar met de niveaus van het Arc5-M1 en het ongewijzigd Arc5a proteïne. Ook het laatste geïntroduceerde Arc5a-M12I2 eiwit met 7 methionine substituties en een methionine-rijke insertie met 5 additionele methionines, werd in grote hoeveelheden geproduceerd in transgene TB1 zaden. Alhoewel een gedeelte van dit gewijzigde eiwit afgebroken werd, waren de bekomen accumulatieniveaus voor het volledige eiwit nog steeds te vergelijken met de niveaus van het ongemodificeerde en de andere gemodificeerde Arc5a eiwitten.

In een pilootexperiment werd de aanwijzing gegeven dat het hoge accumulatieniveau van het Arc5a-M12 eiwit in *P. acutifolius* TB1 zaden kan resulteren in een verhoogd methioninegehalte. Er wordt inderdaad een significante stijging van het methioninegehalte verwacht door accumulatie van een gemodificeerd eiwit met 7 additionele methionines en zeker met het Arc5a-M12I2 met 12 methionines wanneer gelijkaardige accumulatieniveaus behaald worden. Theoretische berekeningen toonden aan dat het methioninegehalte van *Phaseolus* bonen 40 % en 80 % kan stijgen wanneer respectievelijk het Arc5a-M12 en Arc5a-M12I2 20 % van het totale zaadeiwitgehalte uitmaakt.

Mocht uit de analyses blijken dat het Arc5a-M12I2 eiwit niet hoog accumuleert in de getransformeerde *P. acutifolius* zaden of dat sommige methionine residu's verloren zijn gegaan door splitsing of partiële degradatie van het gemodificeerde eiwit, dan is het nog steeds mogelijk om een gemodificeerd Arc5a eiwit te creëren dat 10 methionines bevat door gebruik te maken van de minst risicovolle substituties van groep M3 en M4 (zoals Val94, Ile138 en Leu219) en deze te combineren met de substituties van groep M1 en M2. Wanneer dit gemodificeerd eiwit een accumulatieniveau van 20 % van het totale zaadeiwitgehalte kan halen, is het mogelijk om hiermee het methioninegehalte van *Phaseolus* bonen te verhogen tot

boven de FAO referentiewaarde voor zwavelhoudende aminozuren, ervan uitgaande dat het cysteïnegehalte onveranderd blijft.

Gezien er niet geweten is of er voldoende methionine beschikbaar is in de *Phaseolus* zaden voor de biosynthese van methioninerijke eiwitten, werd er gedacht aan een synergetische aanpak waarbij niet alleen de vraag naar methionine, gecreëerd door accumulatie van een methioninerijk eiwit, maar ook de toevoer van vrij methionine wordt verhoogd. Zoals gezien in hoofdstuk 1, is de methioninebiosynthese zeer complex waarbij veel belangrijke sleutelenzymen en tussenproducten een belangrijke rol spelen. Om het vrij methioninegehalte te verhogen kunnen sommige van de enzymatische stappen in het begin of op een vertakking van de aspartaat biosyntheseweg gewijzigd worden (zie fig. 1.3).

Een eerste enzyme dat aangepakt kan worden is het aspartaatkinase enzyme. Dit enzyme wordt in normale omstandigheden geremd door terugkoppeling van de aminozuren threonine and lysine, eindproducten van de aspartaat biosyntheseweg. Zaadspecifieke expressie van een mutant aspartaatkinase gen uit *Escherichia coli* dat ongevoelig is voor feedback inhibitie, leidde tot een verhoging van het vrije threoninegehalte (16 keer) en methioninegehalte (3 keer) in transgene tabakszaden (Karchi et al., 1993). Ook in *Vicia narbonensis* zaden werd een methioninegehalte bekomen dat 2 tot 3 keer hoger was door expressie van hetzelfde mutante gen (Pickardt et al., 1998). Wij wilden graag hetzelfde mutant aspartaatkinase gen tot expressie brengen in *P. acutifolius* samen met het gemodificeerd Arc5a-M12 eiwit door co-transformatie met twee *Agrobacterium* stammen. Jammer genoeg werden er geen co-transformanten bekomen (zie hoofdstuk 5).

Er zijn nog heel wat andere enzymen die beschouwd kunnen worden. Cystathionine *γ*-synthase (CGS), het eerste enzyme van *de novo* methionine biosynthese, blijkt een belangrijke rol te spelen in methionine synthese. In *Arabidopsis thaliana* zorgt de aanwezigheid van een grote hoeveelheid methionine voor destabilisatie van het CGS mRNA wat leidt tot verminderde CGS activiteit (Chiba et al., 1999). Er kon een *Arabidopsis* mutant worden geïsoleerd die een mutatie bezit verantwoordelijk voor verhoogde stabiliteit van het CGS mRNA (Inaba et al., 1994). Deze mutant accumuleert tot 40 keer meer vrij methionine in de bladeren en 5 tot 8 keer meer in de onrijpe vruchten. Wanneer het gen voor CGS constitutief tot overexpressie werd gebracht, verhoogde het gehalte aan vrij methionine (tot 20 keer) in de meeste delen van de transgene *A. thaliana* planten, ook in de zaden (Kim et al., 2002). In de aspartaat biosyntheseweg concurreren CGS en threoninesynthase (TS) voor hun gemeenschappelijk substraat *O*-phosphohomoserine (OPH). Hierdoor is het mogelijk om gelijkaardige resultaten te bekomen door de activiteit van TS te verlagen. In aardappel werd dit bekomen door gebruik te maken van de antisense strategie. Dit resulteerde in een verlaagd threoninegehalte en sterk verhoogd (60 tot 200 keer) methioninegehalte in de bladeren terwijl in de aardappelknol het threoninegehalte niet veranderd was maar het methionineniveau wel tot 30 keer steeg (Zeh et al., 2001).

Naast de verhoging van de koolstof toevoer naar de methionine synthese kan ook de zwavel assimilatie en cysteine biosynthese verhoogd worden om de methionine productie te stimuleren aangezien cysteine nodig is als bouwstof in de eerste stap van de methionine synthese. Het enzyme serine acetyltransferase wordt gezien als belangrijkste regulatorische element in de cysteine biosynthese (Saito, 1999). Wanneer een bacterieel gen voor SAT constitutief tot overexpressie werd gebracht in tabaks- en aardappelplanten, leidde dit tot een cysteine inhoud van de bladeren die 2 tot 3 keer hoger was (Blaszczyk et al., 1999; Harms et al., 2000). Een ander belangrijk punt is de beschikbaarheid van zwavel. Experimenten uitgevoerd door Tabe en Droux (2002) toonden aan dat opname van sulfaat in de cotylen van transgene lupine zaden de limiterende factor was voor zwavel accumulatie in deze transgene zaden getransformeerd met het methioninerijke 2S albumine van zonnebloem. Recent werd aangetoond in *Arabidopsis* dat een bepaalde sulfaat transporter betrokken is bij de aanvoer van zwavel naar zaden (Awazuhara et al., 2001). Verhoging van de activiteit van dergelijke sulfaat transporters, kan bijdragen tot het verhogen van het methioninegehalte van zaden.

Methionine wordt ook verder omgezet naar S-adenosylmethionine (SAM) en S-methylmethionine (SMM). Om het vrije methionine gehalte van zaden te verhogen kan het overwogen worden om deze omzettingen te verminderen. Co-suppressie van SAM synthetase, het enzyme dat verantwoordelijk is voor de productie van SAM, zorgde voor een 440 keer hoger vrij methionine gehalte in de bladeren van tabaksplanten. Deze hoge methionine niveaus veroorzaakten echter morfologische veranderingen en een onaangename geur (Boerjan et al., 1994).

Uiteindelijk blijkt dat de metabolische biosynthesewegen in planten veel subtieler en ingewikkelder in elkaar zitten dan eerst gedacht werd. Interacties tussen de verschillende organen van de plant en tussen de verschillende compartimenten binnen in de plant cel zorgen voor een ongelofelijke complexiteit. Bovendien beïnvloeden de verschillende biosynthesewegen elkaar door stimulatie of inhibitie van bepaalde tussenstappen. Het wijzigen van één element in deze complexe eenheid kan leiden tot veel veranderingen met serieuze consequenties. Alvorens de biosyntheseweg van methionine te wijzigen, zou het verstandig zijn om eerst te proberen de volledige regulatie ervan op te helderen en te begrijpen.

A

Acosta Gallegos JA Quintero C Vargas J Toro O Tohme J Cardona C (1998) A new variant of arcelin in wild common bean *Phaseolus vulgaris* L. from Southern Mexico. Genet Resour Crop Evol 45: 235-242

Allavena A Rossetti L (1986) Micropropagation of bean (*Phaseolus vulgaris* L.); effects of genetic, epigenetic and environmental factors. Sci Hortic 30: 37–46

Altenbach SB Pearson KW Leung FW Sun SSM (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant Mol Biol 8: 239-250

Altenbach SB Pearson KW Meeker G Staraci LC Sun SSM (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol Biol 13: 513-522

Altenbach SB Kuo C Staraci LC Pearson KW Wainwright C Georgescu A Townsend J (1992) Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol Biol 18: 235-245

Ampe C Van Damme J Castro LAB Sampaio MJA Van Montagu M Vandekerckhove J (1986) The amino acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* HBK). Eur J Biochem 159: 597-604

Anderson JW Fitzgerald MA (2001) Physiological and metabolic origin of sulphur for the synthesis of seed storage proteins. J Plant Physiol 158: 447-456

Anthony J Brown W Buhr D Ronhovde G Genovesi D Lane T Yingling R Aves K Rosato M Anderson P (1997) Transgenic maize with elevated 10 kD zien and methionine. In: Sulphur Metabolism in Higher Plants: Molecular, Ecophysiological and Nutritional Aspects (Cram WJ De Kok LJ Stulen I Brunold C Rennenberg H eds), Backhuys Publishers, Leiden, pp 295-297

Aragão FJL Barros LMG Brasileiro ACM Ribeiro SG Smith FD Sanford JC Faria JC Rech EL (1996) Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardement. Theor Appl Genet 93: 142-150

Aragão FJL Barros LMG de Sousa MV Grossi de Sa MF Almeida ERP Gander ER Rech EL (1999) Expression of a methionine-rich storage albumin from Brazil nut (*Bertholletia excelsa*) in transgenic bean plants (*Phaseolus vulgaris* L.). Genet Mol Biol 22: 445-449

Atkins CA Flinn AM (1978) Carbon dioxide fixation in the carbon economy of developing seeds of *Lupinus albus* (L.). Plant Physiol 62: 486-490

Awazuhara M Takahashi H Watanabe-Takahshi A Hayash H Fujiwara T Saito K (2001) Function of the sulfate transporter Sultr2;1 in seeds of *Arabidopsis thaliana*. In: Plant Nutrition Food Security and Sustainability of Agro-ecosystems (Horst WJ Schenk MK Burkert A eds), Kluwer Academic Publishers, Dordrecht, pp 38-39

Azevedo RA Smith RJ Lea PJ (1992) Aspartate kinase regulation in maize - evidence for co-purification of threonine-sensitive aspartate kinase and homoserine dehydrogenase. Phytochemistry 31: 3731-3734

Azevedo RA Arruda P Turner WL Lea PJ (1997) The biosynthesis and metabolism of the aspartate derived amino acids in higher plants. Phytochemistry 46: 395-419

B

Bagga S Adams H Kemp JD Sengupta-Gopalan C (1995) Accumulation of 15-kilodalton zein in novel protein bodies in transgenic tobacco. Plant Physiol 107: 13-23

Banerjee R Mande SC Ganesh V Das K Dhanaraj V Mahanta SK Suguna K Surolia A Vijayan M (1994) Crystal structure of peanut lectin A protein with an unusual quaternary structure. Proc Natl Acad Sci USA 91: 227-231

Bartlem D Lambein I Okamoto T (2000) Mutation in the threonine synthase gene results in an over-accumulation of soluble methionine in Arabidopsis. Plant Physiol 123: 101-110

Baum HJ Madison JT Thompson JF (1983) Feedback inhibition of homoserine kinase from radish leaves. Phytochemistry 22: 2409-2412

Bäumlein H Wobus U Pustell J Kafatos FC (1986) The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res 14: 2707-2720

Beach LR Ballo B (1992) Enhancing the nutritional value of seed crops. In: Biosynthesis and Molecular Regulation of Amino Acids in Plants (Singh BK Flores HE Shannon JC eds), American Society of Plant Physiologists, Rockville, MD, pp 229-238

Beachy RN Chen Z-L Horsch RB Rogers SG Hoffmann NJ Fraley RT (1985) Accumulation and assembly of soybean ß-conglycinin in seeds of transformed petunia plants. EMBO J 4: 3047-3053

Bechtold N Ellis J Pelletier G (1993) In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. Comptes Rendus Acad Sci Sci Ser III-Sci Vie-Life Sci 316: 1194-1199

Becker J Vogel T Iqbal J Nagl W (1994) *Agrobacterium*-mediated transformation of *Phaseolus vulgaris:* Adaptation of some conditions. Ann Rep Bean Improv Coop 37: 127-128

Becker JW Reeke GN Jr Wang JL Cunningham BA Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A. III: Structure of the monomer and its interactions with metals and saccharides. Biol Chem 250: 1513-1524

Birch RG (1997) Plant transformation: problems and strategies for practical application. Annu Rev Plant Physiol Plant Mol Biol 48: 297-326

Blaszczyk A Brodzik R Sirko A (1999) Increased resistance to oxidative stress in transgenic tobacco plants overexpressing bacterial serine acetyltransferase. Plant J 20: 237-243

Blaygrove RJ Gillespie JM Randall PJ (1976) Effect of nitrogen source and concentration on the free amino acid composition of developing wheat grains. Aust J Plant Physiol 17: 199-206

Bliss FA (1990) Genetic alteration of legume seed proteins. HortScience 25: 1517-1520

Boase MR Bradley JM Borst NK (1998) An improved method for transformation of regal pelargonium (*Pelargonium* x *domesticum* Dubonnet) by *Agrobacterium tumefaciens*. Plant Sci 139: 59-69

Boerjan W Bauw G Van Montagu M Inzé D (1994) Distinct phenotypes generated by overexpression and suppression of *S*-adenosyl-L-methionine synthetase reveal developmental patterns of gene silencing in tobacco. Plant Cell 6: 1401-1414

Bogdanova N Hell R (1997) Cysteine synthesis in plants: Protein-protein interactions of serine acetyltransferase from *Arabidopsis thaliana*. Plant J 11: 251–262

Bollini R Chrispeels MJ (1978) Characterization and subcellular localization of vicilin and phytohemagglutinin, the two major reserve proteins of *Phaseolus vulgaris* L. Planta 142: 291-298

Bollini R Van der Wilden W Chrispeels MJ (1982) A precursor of the reserve-protein, phaseolin, is transiently associated with the endoplasmic reticulum of developing *Phaseolus vulgaris* cotyledons. Physiol Plant 55: 82-92

Bork C Schwenn JD Hell R (1998) Isolation and characterization of a gene for assimilatory sulfite reductase from *Arabidopsis thaliana*. Gene 212: 147–153

Boston RS Viitanen PV Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32: 191-222

Bourgis F Roje S Nuccio ML Fisher DB Tarczynski MC Li C Herschbach C Rennenberg H Pimenta MJ Shen T-L Gage DA Hanson AD (1999) S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. Plant Cell 11: 1485–1497

Bourne Y Roussel A Frey M Rouge P Fontecilla-Camps JC Cambillau C (1990) Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: Fine specificity of the monosaccharide-binding site. Proteins 8: 365-376

Bourne Y Mazurier J Legrand D Rouge P Montreuil J Spik G Cambillau C (1994) Structures of a legume lectin complexed with the human lactotransferrin N2 fragment and with an isolated biantennary glycopeptide: Role of the fucose moiety. Structure 15: 209-219

Bressani R Elias LG (1968) Processed vegetable protein mixtures for human consumption in developing countries. Adv Food Res 16: 1-103

Breyne P Van Montagu M Gheysen G (1994) The role of scaffold attachment regions in the structural and functional organization of plant chromatin. Transgenic Res 3: 195-202

Bright S Shewry PR (1983) Improvement of protein quality in cereals. Crit Rev Plant Sci 1: 49-93

Bright SWJ Miflin BJ Rognes SE (1982) Threonine accumulation in seed of a barley mutant with an altered aspartate kinase. Biochem Genet 20: 229-243

Brown JWS Ma Y Bliss FA Hall TC (1981a) Genetic variation in the subunits of globulin-1 storage protein of French bean. Theor Appl Genet 59: 83-88

Brown JWS Osborn TC Bliss FA Hall TC (1981b) Genetic variation in the subunits of globulin-2 and albumin seed proteins of French bean. Theor Appl Genet 60: 245-250

Brunold C Rennenberg H (1997) Regulation of sulfur metabolism in plants: First molecular approaches. Prog Bot 58: 164–186

Bryan JK (1980) Synthesis of the aspartate family and branched-chain amino acids. In: Amino Acids and Derivatives (The Biochemistry of Plants, A Comprehensive Treatise, Vol 5) (Miflin BJ ed), Academic Press, New York, pp 403-452

Bryan JK (1990) Advances in the biochemistry of amino acid biosynthesis. In: Intermediary Nitrogen Metabolism (The Biochemistry of Plants: A Comprehensive Treatise, Vol 16) (Miflin BJ Lea PJ eds), Academic Press, New York, pp 161–196

Bustos MM Kalkan FA VandenBosch KA Hall TC (1991) Differential accumulation of four phaseolin glycoforms in transgenic tobacco. Plant Mol Biol 16: 381-395

C

Cangelosi GA Ankenbauer RG Nester EW (1990) Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. Proc Natl Acad Sci USA 87: 6708-6712

Cao X Liu Q Rowland LJ Hammerschlag FA (1998) GUS expression in blueberry (*Vaccinium* spp): Factors influencing *Agrobacterium*-mediated gene transfer efficiency. Plant Cell Rep 18: 266-270

Casazza AP Basner A Höfgen R Hesse H (2000) Expression of threonine synthase from *Solanum tuberosum* L. is not metabolically regulated by photosynthesis-related signals or by nitrogenous compounds. Plant Sci 157: 43-50

Casey R Domoney C Ellis N (1986) Legume storage proteins and their genes. Oxford Surv Plant Mol Cell Biol 3: 2-95

Cattoir-Reynaerts A Degryse E Verbruggen I Jacobs M (1983) Selection and characterization of carrot embryoid cultures resistant to inhibition by lysine and threonine. Biochem Physiol Pflanz 178: 89-90

Ceriotti A Pedrazzini E Bielli A Giovinazzo G Bollini R Vitale A (1995) Assembly and intracellular transport of phaseolin, the major storage protein of *Phaseolus vulgaris* L. J Plant Physiol 145: 648–653

Chandler PM Higgins TJV Randall PJ Spencer D (1983) Regulation of legumin levels in developing pea seeds under conditions of sulfur deficiency rates of legumin synthesis and levels of legumin mRNA. Plant Physiol 71: 47-54

Chandler PM Spencer D Randall PJ Higgins TJV (1984) Influence of sulfur nutrition on developmental patterns of some major pea seed proteins and their mRNAs. Plant Physiol 75: 651-657

Chappell J Van der Wilden W Chrispeels MJ (1980) The biosynthesis of ribonuclease and its accumulation in protein bodies in the cotyledons of Mung beans. Dev Biol 76: 115–125

Chiba Y Ishikawa M Kijima F Tyson RH Kim J Yamamoto A Mambara E Leustek T Wallsgrove RM Naito S (1999) Evidence for autoregulation of cystathionine γ-synthase mRNA stability in *Arabidopsis*. Science 286: 1371–1374

Choi YE Yang DC Kusano T Sano H (2001) Rapid and efficient *Agrobacterium*-mediated transformation of *Panax ginseng* by plasmolyzing pre-treatment of cotyledons. Plant Cell Rep 20: 616-621

Chrispeels MJ Higgins TJV Craig S Spencer D (1982a) Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. J Cell Biol 93: 5–14

Chrispeels MJ Higgins JV Spencer D (1982b) Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. J Cell Biol 93: 306–313

Chrispeels MJ (1991) Sorting of proteins in the secretory system. Annu Rev Plant Physiol Plant Mol Biol 42: 21-53

Chrispeels MJ Raikhel NV (1991) Lectins, lectin genes, and their role in plant defense. Plant Cell 3: 1-9

Christiansen P Gibson JM Moore A Pedersen C Tabe L Larkin PJ (2000) Transgenic *Trifolium repens* with foliage accumulating the high sulphur protein sunflower seed albumin. Transgenic Res 9: 103-113

Clough SJ Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16: 735-743

Cohen GN Saint-Girons I (1987) Biosynthesis of threonine lysine and methionine in *Escherichia coli* and *Salmonella typhimurium*. In: Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, Vol. I (Neidhardt FC Ingraham JL Low KB Magasanik B Schaechter M Umbarger HE eds) American Society for Microbiology, Washington, pp 429-444

Collaborative Computational Project, Number 4 (1994). The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr Sect D 50: 760-763

Conceição AD Van Vliet A Krebbers E (1994) Unexpectedly higher expression levels of a chimeric 2S albumin seed protein transgene from a tandem array construct. Plant Mol Biol 26: 1001-1005

Craciun A Jacobs M Vauterin M (2000) Arabidopsis loss-of-function mutant in the lysine pathway points out complex regulation mechanisms. FEBS Lett 487: 234-238

Cruz de Carvalho MH Van Le B Zuily-Fodil Y Pham Thi AT Tran Thanh Van K (2000) Efficient whole plant regeneration of common bean (*Phaseolus vulgaris* L.) using thin-cell-layer culture and silver nitrate. Plant Sci 159: 223-232

Curien G Dumas R Ravanel S (1996) Characterization of an *Arabidopsis thaliana* cDNA encoding an S-adenosylmethionine-sensitive threonine synthase from higher plants. FEBS Lett 390: 85-90

Curien G Job D Douce R Dumas R (1998) Allosteric activation of *Arabidopsis* threonine synthase by S-adenosylmethionine. Biochemistry 37: 13212-13221

D

Datko AH Giovanelli J Mudd SH (1974) Homocysteine biosynthesis in green plants: O-phosphohomoserine is the physiological substrate for cystathionine *ß*-synthase. J Biol Chem 249: 1139–1155

De Block M Botterman J Vandewiele M Dockx J Thoen C Gosselé V Movva R Thompson C Van Montagu M Leemans J (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J 6: 2513-2518

De Block M Debrouwer D (1991) Two T-DNA's co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. Theor Appl Genet 82: 257-263

De Bondt A Eggermont K Druart P De Vil M Goderis I Vanderleyden J Broekaert WF (1994) *Agrobacterium*-mediated transformation of apple (*Malus* × *domestica* Borkh): an assessment of factors affecting gene transfer efficiency during early transformation steps. Plant Cell Rep 13: 587-593

De Buck S Jacobs A Van Montagu M Depicker A (1998) *Agrobacterium tumefaciens* transformation and cotransformation frequencies of *Arabidopsis thaliana* root explants and tobacco protoplasts. Mol Plant-Microbe Interact 11: 449-457

De Clercq A Vandewiele M Van Damme J Guerche P Van Montagu M Vandekerckhove J Krebbers E (1990) Stable accumulation of modified 2S albumin seed storage proteins with higher methionine contents in transgenic plants. Plant Physiol 94: 970-979

De Clercq J Zambre M Van Montagu M Dillen W Angenon G (2002) An optimized Agrobacterium-mediated transformation procedure for *Phaseolus acutifolius* A. Gray. Plant Cell Rep: in press

Dean C Tamaki S Dunsmuir P Favreau M Katayama C Dooner H Bedbrook J (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. Nucleic Acids Res 14: 2229-2240

Deblaere R Bytebier B De Greve H Deboeck F Schell J Van Montagu M Leemans J (1985) Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. Nucleic Acids Res 13: 4777-4788

Deblaere R Reynaerts A Höfte H Hernalsteens J-P Leemans J Van Montagu M (1987) Vectors for cloning in plant cells. *In* R Wu L Grossman eds Recombinant DNA: Part D. Methods in Enzymology Vol 153. Academic Press New York pp 277-292

Delbaere LTJ Vandonselaar M Prasad L Quail JW Nikrad PV Pearlstone JR Carpenter MR Smillie LB Spohr U Lemieux RU (1989) Trans ACA 25: 65-76

Dellaporta SL Wood J Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Reporter 1: 19-21

de Pater S Pham K Klitsie I Kijne J (1996) The 22 bp W1 element in the pea lectin promoter is necessary and as a multimer sufficient for high gene expression in tobacco seeds. Plant Mol Biol 32: 515-523

Denecke et al.(1989) Quantitative Analysis of Transiently expressed Genes in Plant Cells. Methods Mol Cell Biol 1:19-27

Denis M Van Vliet A Leyns F Krebbers E Renard M (1995) Field evaluation in transgenic *Brassica napus* lines carrying a seed-specific chimeric 2S albumin gene. Plant Breed 111: 97-107

Dessen A Gupta D Sabesan S Brewer CF Sacchettini JC (1995) X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. Biochemistry 34: 4933-4942

Dhaese P De Greve H Gielen J Seurinck J Van Montagu M Schell J (1983) Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using *Agrobacterium* T-DNA genes as models. EMBO J 2: 419-426

D'Halluin K De Block M Denecke J Janssens J Leeman J Reynaerts A Botterman J (1992) The *bar* gene as selectable and screenable marker in plant engineering. Meth Enzymol 216: 415-426

Diaz CL Melchers LS Hooykaas PJJ Lugtenberg BJJ Kijne JW (1989) Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. Nature 338: 579-581

Diedrick TJ Frisch DA Gengenbach BG (1990) Tissue culture isolation of a second mutant locus for increased threonine accumulation in maize. Theor Appl Genet 79: 209-215

Dillen W De Clercq J Van Montagu M Angenon G (1996) Plant regeneration from callus in a range of *Phaseolus acutifolius* A. Gray genotypes. Plant Sci 118: 81-88

Dillen W De Clercq J Goossens A Van Montagu M Angenon G (1997a) *Agrobacterium*-mediated transformation of *Phaseolus acutifolius* A. Gray. Theor Appl Genet 94: 151-158

Dillen W De Clercq J Kapila J Zambre M Van Montagu M Angenon G (1997b) The effect of temperature on *Agrobacterium tumefaciens*-mediated gene transfer to plants. Plant J 12: 1459-1463

Dinkins RD Reddy MSS Meurer CA Yan B Trick H Thibaud-Nissen F Finer JJ Parrott WA Collins GB (2001) Increased sulfur amino acids in soybean plants overexpressing the maize 15 kDa zein protein. In Vitro Cell Dev Biol Plant 37: 742-747

Donaldson PA Simmonds DH (2000) Susceptibility to *Agrobacterium tumefaciens* and cotyledonary node transformation in short-season soybean. Plant Cell Rep 19: 478-484

Dotson SB Frisch DA Somers DA Gengenbach BG (1990) Lysine-insensitive aspartate kinase in two threonine-overproducing mutants of maize. Planta 182: 546-552

Doyle JJ Schulers MA Godette WD Zenger V Beachy RN (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. J Biol Chem 261: 9228-9238

Droux M Ravanel S Douce R (1995) Methionine biosynthesis in higher plants. II: Purification and characterisation of cystathionine-γ-lyase from spinach chloroplasts. Arch Biochem Biophys 316: 585–595

Droux M Ruffet ML Douce R (1998) Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants - Structural and kinetic properties of the free and bound enzymes. Eur J Biochem 255: 235-245

Dudman WF Millerd A (1975) Immunochemical behaviour of legumin and vicilin from *Vicia faba*: Survey of related proteins in Leguminosae subfamily Faboideae. Biochem Syst Ecol 3: 25-33

Durante M Bernardi R Lupi MC Pini S (1989) *Phaseolus coccineus* storage proteins II: Electrophoretic analysis and erythroagglutinating activity in various cultivars. Plant Breed 102: 58-65

Dyer JM Nelson JW Murai N (1993) Strategies for selecting mutation sites for methionine enhancement in the bean seed storage protein phaseolin. J Prot Chem 12: 545-560

Dyer JM Nelson JW Murai N (1995) Extensive modifications for methionine enhancement in the β-barrels do not alter the structural stability of the bean seed storage protein phaseolin. J Prot Chem 14: 665-678

E

Egnin M Mora A Prakash CS (1998) Factors enhancing *Agrobacterium tumefaciens*-mediated gene transfer in peanut (*Arachis hypogaea* L). In Vitro Cell Dev Biol-Plant 34: 310-318

Eichel J González JC Hotze M Matthews RG Schröder J (1995) Vitamin-B12-independent methionine synthase from higher plant (*Catharanthus roseus*): Molecular characterization, regulation, heterologous expression and enzyme properties. Eur J Biochem 230: 1053–1058

Einspahr H Parks EH Suguna K Subramanian E Suddath FL (1986) The crystal structure of pea lectin at 30-Å resolution. J Biol Chem 261 16518-16527

Ellis JR Shirsat AH Hepher A Yarwood JN Gatehouse JA Croy RRD Boulter D (1988) Tissue-specific expression of a pea legumin gene in seeds of *Nicotiana plumbaginifolia*. Plant Mol Biol 10: 203-214

Enneking D (1995) The toxicity of Vicia species and their utilisation as grain legumes. Phd thesis, Clima, Adelaide

F

Falco SC Guida T Locke M Mauvais J Sandres C Ward RT Webber P (1995) Transgenic canola and soybean seeds with increased lysine. Bio/Technology 13: 577-582

FAO (1970) Amino-Acid content of foods and biological data on proteins. FAO: Nutritional Studies 24: 50

FAO (1990) FAO Year book in Production vol 44. FAP Statistics Series 99: 61

Frankard V Gislain M Negrutiu I Jacobs M (1991) High threonine producer mutant of *Nicotiana sylvestris*. Theor Appl Genet 82: 273-282

Frankard V Ghislain M Jacobs M (1992) Two feedback-insensitive enzymes of the aspartate pathway in *Nicotiona sylvestris*. Plant Physiol 99: 1285-1293

Frankard V Ghislain M Jacobs M (1997) Molecular characterization of an *Arabidopsis thaliana* cDNA coding for a monofunctional aspartate kinase. Plant Mol Biol 34: 233-42

Franklin CI Trieu TN Gonzales RA Dixon RA (1991) Plant regeneration from seedling explants of green bean (*Phaseolus vulgaris* L.) via organogenesis. Plant Cell Tissue Organ Cult 24: 199-206

Frigerio L de Virgilio M Prada A Faoro F Vitale A (1998) Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. Plant Cell 10: 1031-1042

Frigerio L Foresti O Neuhaus JM Vitale A (2001) The C-terminal tetrapeptide of phaseolin is sufficient to target green fluorescent protein to the vacuole. J Plant Physiol 158: 499-503

Fujiwara T Hirai MY Chino M Komada Y Naito S (1992) Effects of sulfur nutrition on expression of the soybean seed storage protein genes in transgenic petunia. Plant Physiol 99: 263-268

Fuller MF, Mennie I, Crofts RMJ (1989) The optimum dietary amino acid pattern for growing pigs. Br J Nutr 62: 255-267

G

Gabière B Ravanel S Droux M Douce R Job D (2000) Mechanisms to account for maintenance of the soluble methionine pool in transgenic *Arabidopsis* plants expressing antisense cystathionine γ-synthase cDNA. Comptes Rendus Acad Sci Sci Ser III-Sci Vie-Life Sci 323: 841-851

Galili G (1995) Regulation of lysine and threonine synthesis. Plant Cell 7: 899-905

Galili G Herman EM (1997) Protein bodies: Storage vacuoles in seeds. Adv Bot Res 25: 113–140

Galili G Larkins BA (1999) Enhancing the content of the essential amino acids lysine and threonine in plants. In: Plant Amino Acids: Biochemistry and Biotechnology (Singh BK ed), Academic Press, New York, pp 487-507

Galili G Höfgen R (2002) Metabolic engineering of amino acids and storage proteins in plants. Metab Eng 4: 3-11

Gayler KR Sykes GE (1985) Effects of nutritional stress on the strorage proteins of soybeans. Plant Physiol 78: 582-585

Gelvin SB (2000) *Agrobacterium* and plant genes involved in T-DNA transfer and integration. Annu Rev Plant Physiol Plant Mol Biol 51: 223-256

George A de Lumen BO (1991) A novel methionine rich protein in soybean seed: identification amino acid composition and N-terminus sequence. J Agric Food Chem 39: 224-270

Gepts P Bliss FA (1984) Enhanced available methionine concentration associated with higher phaseolin levels in common bean seeds. Theor Appl Genet 69: 47-53

Gheysen G Angenon G Van Montagu M (1998) *Agrobacterium*-mediated plant transformation: A scientifically intriguing story with significant applications. In: Transgenic Plant Research (Lindsey K ed) Harwood Academic Publishers, Amsterdam, pp 1-33

Ghislain M Frankard V Vandenbossche D Matthews BF Jacobs M (1994) Molecular analysis of the aspartate kinase-homoserine dehydrogenase gene from *Arabidopsis thaliana*. Plant Mol Biol 24: 835-851

Giovanelli J Veluthambi K Thompson GA Mudd SH Datko AH (1984) Threonine synthase of *Lemna paucicostata* Hegelm. 6746. Plant Physiol 76: 285-292

Giovanelli J Mudd SH Datko AH (1985) *In vivo* regulation of *de novo* methionine biosynthesis in a higher plant (*Lemna*). Plant Physiol 62: 629-635

Giovinazzo G Greco V Vitale A Bollini R (1997) Bean (*Phaseolus vulgaris* L.) protoplasts as a model system to study the expression and stability of recombinant seed proteins. Plant Cel Rep 16: 705-709

Glover DV Mertz ET (1987) Corn. In: Nutritional Quality of Cereal Grains: Genetic and Agronomic Improvement (Madison O ed) ASA-CSSA-SSA, pp 183-336

Goldberg RB Barker SJ Terez-Grau L (1989) Regulation of gene expression during plant embryogenesis. Cell 56: 149-160

Goossens A Geremia R Bauw G Van Montagu M Angenon G (1994) Isolation and characterisation of arcelin-5 proteins and cDNAs. Eur J Biochem 225: 787-795

Goossens A Ardiles Diaz W De Keyser A Van Montagu M Angenon G (1995) Nucleotide sequence of an arcelin 5-I genomic clone from wild *Phaseolus vulgaris* (Accession No Z50202). Plant Physiol 109: 722

Goossens A (1998) Molecular characterisation of the gene encoding arcelin 5, a seed storage protein from insect resistant wild common beans (*Phaseolus vulgaris*). PhD thesis, Universiteit Gent Belgium

Goossens A Van Montagu M Angenon G (1999b) Co-introduction of an antisense gene for an endogenous seed storage protein can increase expression of a transgene in *Arabidopsis thaliana* seeds. FEBS Lett 459: 160-164

Goossens A Dillen W De Clercq J Van Montagu M Angenon G (1999a) The *arcelin-5* gene of *Phaseolus vulgaris* directs high seed-specific expression in transgenic *Phaseolus acutifolius* and Arabidopsis plants. Plant Physiol 120: 1095-1104

Goossens A Quintero C Dillen W De Rycke R Valor JF De Clercq J Van Montagu M Cardona C Angenon G (2000) Analysis of bruchid resistance in the wild common bean accession G02771: no evidence for insecticidal activity of arcelin 5. J Exp Bot 51: 1229-1236

Grimsley N Hohn B Hohn T Walden R (1986) "Agroinfection", an alternative route for viral infection of plants by using the Ti plasmid. Proc Natl Acad Sci USA 83: 3282-3286

Grossman A Takahashi H (2001) Macronutrient utilization by photosynthetic eukaryotes and the fabric of interactions. Annu Rev Plant Physiol Plant Mol Biol 52: 163-210

Guerche P De Almeida ERP Schwarztein MA Gander E Krebbers E Pelletier G (1990) Expression of the 2S albumin from *Bertholletia excelsa* in *Brassica napus*. Mol Gen Genet 221: 306-314

Guivarc'h A Caissard J-C Brown S Marie D Dewitte W Van Onckelen H Chriqui D (1993) Localization of target cells and improvement of *Agrobacterium*-mediated transformation efficiency by direct acetosyringone pretreatment of carrot root discs. Protoplasma 174: 10-18

Guo G Maiwald F Lorenzen P Steinbiss H-H (1998) Factors influencing T-DNA transfer into wheat and barley cells by *Agrobacterium tumefaciens*. Cereal Res Comm 26: 15-22

H

Hamelryck TW Poortmans F Goossens A Angenon G Wyns L Loris R (1996a) Crystallographic structure of arcelin-5, a lectin-like defense protein from *Phaseolus vulgaris*. J Biol Chem 271: 32796-32802

Hamelryck TW Dao-Thi M Poortmans F Chrispeels MJ Wyns L Loris R (1996b) The crystallographic structure of phytohemagglutinin-L. J Biol Chem 271: 20479-20485

Hanson AD Gage DA Shacher-Hill Y (2000) Plant one-carbon metabolism and its engineering. Trends Plant Sci 5: 206-213

Hara-Nishimura I Takeuchi Y Inoue K Nishimura M (1993) Vesicle transport and processing of the precursor to 2S albumin in pumpkin. Plant J 4: 793–800

Hara-Nishimura I Shimada T Hiraiwa N Nishimura M (1995) Vacuolar processing enzyme responsible for maturation of seed proteins. J Plant Physiol 145: 632–640

Hara-Nishimura I Shimada T Hatano K Takeuchi Y Nishimura M (1998) Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. Plant Cell 10: 825–836

Harms K von Ballmoos P Brunold C Höfgen R Hesse H (2000) Expression of a bacterial serine acetyltransferase in transgenic potato plants leads to increased levels of cysteine and glutathione. Plant J 22: 335–343

Hartweck LM Vogelzang RD Osborn TC (1991) Characterization and comparison of arcelin seed protein variants from common bean. Plant Physiol 97: 204-211

Hatzfeld Y Cathala N Grignon C Davidian JC (1998) Effect of ATP sulfurylase overexpression in Bright Yellow 2 tobacco cells. Plant Physiol 116: 1307-1313

Hell R (1997) Molecular physiology of plant sulfur metabolism. Planta 202: 138–148

Heremans B Jacobs M (1995) Threonine accumulation in a mutant of *Arabidopsis thaliana* (L.) Heynh. with an altered aspartate kinase. J Plant Physiol 146: 249-257

Heremans B Jacobs M (1997) A mutant of *Arabidopsis thaliana* (L.) Heynh. with modified control of aspartate kinase by threonine. Biochem Genet 35: 139-153

Herman EM Chrispeels MJ (1980) Phospholipase D and phosphatidic acid phosphatase: Acid hydrolases involved in phospholipid catabolism in mung bean cotyledons. Plant Physiol 66**:** 1001–1007

Herman EM Larkins BA (1999) Protein storage bodies and vacuoles. Plant Cell 11: 603-613

Hesse H Kreft O Maimann S Zeh M Willmitzer L Höfgen R (2001) Approaches towards understanding methionine biosynthesis in higher plants. Amino Acids 20: 281-289

Hibberd KA Walter T Green CE Gengenbach BG (1980) Selection and characterization of feedback-insensitive tissue culture of maize. Planta 148: 183-187

Hiei Y Ohta S Komari T Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6: 271-282

Hiei Y Komari T Kubo T (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. Plant Mol Biol 35: 205-218

Higgins TJV (1984) Synthesis and regulation of major proteins in seeds. Annu Rev Plant Physiol 35: 191-221

Higgins TJV Chandler PM Randall PJ Spencer D Beach LR Blagrove RJ Kortt AA Inglis AS (1986) Gene structure, protein structure and regulation of the synthesis of a sulfur-rich protein in pea seeds. J Biol Chem 261: 11124-11130

Higgins TJV Newbigin EJ Spencer D Llewellyn DJ Craig S (1988) The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. Plant Mol Biol 11: 683-695

Hirai MY Fujiwara T Goto K Komada Y Chino M Naito S (1994) Differential regulation of soybean seed storage protein gene promoter-GUS fusions by exogenously applied methionine in transgenic *Arabidopsis thaliana*. Plant Cell Physiol 35: 927-934

Hirai MY Fujiwara T Chino M Naito S (1995) Effects of sulfate concentrations on the expression of a soybean seed storage protein gene and its reversibility in transgenic *Arabidopsis thaliana*. Plant Cell Physiol 36: 1131-1339

Hoffman LM Donaldson DD Bookland R Rashka K Herman EM (1987) Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seeds. EMBO J 6: 3213-3221

Hoffman LM Donaldson DD Herman EM (1988) A modified storage protein is synthesized, processed and degraded in the seeds of transgenic plants. Plant Mol Biol 11: 717-729

Höfgen R Kreft O Willmitzer L Hesse H (2001) Manipulation of thiol contents in plants. Amino Acids 20: 291-299

Hohl I Robinson DG Chrispeels MJ Hinz G (1996) Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. J Cell Sci 109: 2539–2550

Holmes-Davis R Comai L (1998) Nuclear matrix attachment regions and plant gene expression. Trends Plant Sci 3: 91-97

Holowach LP Thompson JF Madison JT (1984) Effects of exogenous methionine on storage protein composition of soybean cotyledons cultured *in vitro*. Plant Physiol 74: 576-583

Holsters M Silva B Van Vliet F Genetello G De Block M Dhaese P Depicker A Inzé D Engler G Villarroel R Van Montagu M Schell J (1980) The functional organization of the nopaline *A. tumefaciens* plasmid pTiC58. Plasmid 3: 212-230

Hood EE Helmer GL Fraley RT Chilton M-D (1986) The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. J Bacteriol 168: 1291-1301

Horta ACG Sodek L (1997) Free amino acid and storage protein composition of soybean fruit explants and isolated cotyledons cultured with and without methionine. Ann Bot 79: 547-552

Hubbard SJ Thornton JM (1993) 'NACCESS', Computer Program. Department of Biochemistry and Molecular Biology, University College London.

I

Igasaki T Mohri T Ichikawa H Shinohara K (2000) *Agrobacterium tumefaciens*-mediated transformation of *Robinia pseudoacacia*. Plant Cell Rep 19: 448-453

Imsande J (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. Crop Sci 41: 510-515

Inaba K Fujiwara T Chino M Komeda Y Naito S (1994) Isolation of an *Arabidopsis thaliana* mutant, *mto1*, that overaccumulates soluble methionine. Temporal and spatial patterns of soluble methionine accumulation. Plant Physiol 104: 881-887

Ingelbrecht ILW Herman LMF Dekeyser RA Van Montagu MC Depicker AG (1989) Different 3' end regions strongly influence the level of gene expression in plant cells. Plant Cell 1: 671-680

Ishida Y Saito H Ohta S Hiei Y Komari T Kumashiro T (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. Nature Biotechnol 14: 745-750

J

Jacobs M Negrutiu I Dirks R Cammaerts D (1987) Selection programs for isolation and analysis of mutants in plant cell cultures. In: Plant Biology vol 3: Plant tissue and cell culture (Green CE Somers DA Hackett WP Biesboer DD eds), AR Liss, New York, pp 243-320

Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5: 387-405

Jia S-R Yang M-Z Ott R Chua N-H (1989) High frequency transformation of *Kalanchoe laciniata*. Plant Cell Rep 8: 336-340

Jones TA Zou JY Cowan SW Kjeldgaard (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr Sect A 47: 110-119

Joshi CP (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucleic Acids Res 15: 9627-9639

Jung R Martino-Catt S Townsend J Beach L (1997) Expression of a sulfur-rich protein in soybean seeds causes an altered seed protein composition. Plant Mol Biol Rep Suppl 15:307

Jung R Scott MP Nam YW Beaman TW Bassuner R Saalbach I Muntz K Nielsen NC (1998) The role of proteolysis in the processing and assembly of 11S seed globulins. Plant Cell 10: 343–357

K

Kapila J De Rycke R Van Montagu M Angenon G (1997) An *Agrobacterium*-mediated transient gene expression system for intact leaves. Plant Sci 122: 101-108

Karchi H Shaul O Galili G (1993) Seed-specific expression of a bacterial desensitized aspartate kinase increases the production of seed threonine and methionine in transgenic tabacco. Plant J 3: 721-727

Karchi H Shaul O Galili G (1994) Lysine synthesis and catabolism are coordinately regulated during tobacco seed development. Proc Natl Acad Sci USA 91: 2577-2581

Kartha KK Pahl K Leung NL Mroginski LA (1981) Plant regeneration from meristems of grain legumes: Soybean, cowpea, peanut, chickpea, and bean. Can J Bot 59: 1671-1679

Katsube T Kurisaka N Ogaqa M Maruyama N Ohtsuka R Utsumi S Takaiwa F (1999) Accumulation of soybean glycinin and its assembly with the glutelins in rice. Plant Physiol 120: 1063-1073

Keeler S Maloney C Webber P Patterson C Hirata L Falco S Rice J (1997) Expression of *de novo* high-lysine α-helical coiled-coil proteins may significantly increase the accumulated levels of lysine in mature seeds of transgenic tobacco plants. Plant Mol Biol 34: 15-29

Kelly JD Hefle SL (2000) 2S methionine-rich protein (SSA) from sunflower seed is an IgE-binding protein. Allergy 55: 556-559

Khan MRI Ceriotti A Tabe L Aryan A McNabb W Moore A Craig S Spencer D Higgins TJV (1996) Accumulation of a sulphur-rich seed albumin from sunflower in the leaves of transgenic subterranean clover (*Trifolium subterraneum* L.). Transgenic Res 5: 179-185

Kho CJ de Lumen BO (1988) Identification and isolation of methionine-cysteine rich proteins in soybean seed. Plant Food Hum Nutr 38: 287-296

Kim CS Kamiya S Sato T Utsumi S Kito M (1990) Improvement of nutritional value and functional properties of soybean glycinin by protein engineering. Protein Eng 3: 725-731

Kim J Leustek T (2000) Repression of cystathionine γ-synthase in *Arabidopsis thaliana* produces partial methionine auxotrophy and developmental abnormalties. Plant Sci 151: 9–18

Kim J Lee M Chalam R Martin MN Leustek T Boerjan W (2002) Constitutive overexpression of cystathionine γ-synthase in *Arabidopsis* leads to accumulation of soluble methionine and S-methylmethionine. Plant Physiol 128: 95-107

King SP Badger MR Furbank RT (1998) CO_2 refixation characteristics of developing canola seeds and silique wall. Aust J Plant Physiol 25: 377-386

Kircher M Huthmacher K (1999) Amino acids enhance the nutritional value of feeds. Agr-Food Industry Hi-Tech may/june: 7-11

Kirchgeßner M Windisch W Roth FX (1994) Nova Acta Leopoldina NF 70: 393-412

Kjemtrup S Herman EM Chrispeels M (1994) Correct post-translational modification and stable vacuolar accumulation of phytohemagglutinin engineered to contain multiple methionine residues. Eur J Biochem 226: 385-391

Kohno-Murase J Murase M Ichikawa H Imamura J (1995) Improvement in the quality of seed storage protein by transformation of *Brassica napus* with an antisense gene for cruciferin. Theor Appl Genet 91: 627-631

Koncz C Schell J (1986) The promoter of T_L-DNA gene *5* controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol Gen Genet 204: 383-396

Kondo T Hasegawa H Suzuki M (2000) Transformation and regeneration of garlic (*Allium sativum* L.) by *Agrobacterium*-mediated gene transfer. Plant Cell Rep 19: 989-993

Kopriva S Muheim R Koprivova A Trachsel N Catalano C Suter M Brunold C (1999) Light regulation of assimilatory sulphate reduction in *Arabidopsis thaliana*. Plant J 20: 37–44

Koprivova A Suter M Op den Camp R Brunold C Kopriva S (2000) Regulation of sulfate assimilation by nitrogen in *Arabidopsis thaliana*. Plant Physiol 122: 737-746

Kortt AA Caldwell JB (1990) Molecular weight albumins from sunflower seed - Identification of a methionine-rich albumin. Phytochemistry 29: 2805-2810

Kraulis PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24: 946-950

Kreis M Shewry PR Forde BG Forde J Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surv Plant Mol Cell Biol 2: 253-317

L

Lawrence MC Suzuki E Varghese JN Davis PC Van Donkelaar A Tulloch PA Colman PM (1990) The three-dimensional structure of the seeds storage protein phaseolin at 3 Å resolution. EMBO J 9: 9-15

Lea PJ Miflin BJ (1980) Transport and metabolism of asparagine and other nitrogen compounds within the plant. In: Amino Acids and Derivatives (The Biochemistry of Plants, A Comprehensive Treatise, Vol 5) (Miflin BJ ed), Academic Press, New York, pp 569-607

Leustek T Saito K (1999) Sulfate transport and assimilation in plants. Plant Physiol 120: 637–643

Leustek T Martin MN Bick JA Davies JP (2000) Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. Annu Rev Plant Physiol Plant Mol Biol 51: 141-166

Lioi L Bollini R (1989) Identification of a new arcelin variant in wild bean seeds. Bean Improv Coop 32, 28

Loris R Steyaert J Maes D Lisgarten J Pickersgill R Wyns L (1993) Crystal structure determination and refinement at 23-Å resolution of the lentil lectin. Biochemistry 32: 8772-8781

M

Madison JT Thompson JF (1988) Characterization of soybean tissue culture cell lines resistant to methionine analogs. Plant Cell Rep 7: 473-476

Maimann S Wagner C Kreft O Zeh M Willmitzer L Höfgen R Hesse H (2000) Transgenic potato plants reveal the indispensable role of cystathionine β-lyase in plant growth and development. Plant J 23: 747-758

Maimann S Höfgen R Hesse H (2001) Enhanced cystathionine β-lyase activity in transgenic potato plants does not force metabolite flow towards methionine. Planta 214: 163-170

Malik KA Saxena PK (1992) Regeneration in *Phaseolus vulgaris* L.: High-frequency induction of direct shoot formation in intact seedlings by N^6-benzyl-aminopurine and thidiazuron. Planta 186: 384-389

Martins IS Sondahl MR (1984) Early stages of somatic embryo differentiation from callus cells of bean (*Phaseolus vulgaris* L.) grown in liquid medium. J Plant Physiol 117: 97-103

Matthews B (1999) Lysine threonine and methionine biosynthesis. In: Plant Amino Acids (Singh BK ed) Marcel Dekker, New York, pp. 205–225

Maughan PJ Philip R Cho MJ Widholm JM Vodkin LO (1999) Biolistic transformation, expression, and inheritance of bovine β-casein in soybean (*Glycine max*). In Vitro Cell Dev Biol Plant 35: 334-349

Maximova SN Dandekar AM Guiltinan MJ (1998) Investigation of *Agrobacterium*-mediated transformation of apple using green fluorescent protein: High transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting. Plant Mol Biol 37: 549-559

McClean P Grafton KF (1989) Regeneration of dry bean (*Phaseolus vulgaris* L.) via organogenesis. Plant Sci 60: 117-122

Mertz ET Bates LS Nelson OE (1964) Mutant gene that changes protein composition and incrased lysine content of maize endosperm. Science 145: 279-280

Mettler IJ Beevers H (1979) Isolation and characterization of the protein body membrane of castor beans. Plant Physiol 64: 506–511

Meurer CA Dinkins RD Collins GB (1998) Factors affecting soybean cotyledonary node transformation. Plant Cell Rep 18: 180-186

Minney BHP Gatehouse AMR Dobie P Dendy J Cardona C Gatehouse JA (1990) Biochemical bases of seed resistance to *Zabrotes subfasciatus* (bean weevil) in *Phaseolus vulgaris* (common bean); a mechanism for arcelin toxicity. J Insect Physiol 36: 757-767

Mirkov TE Wahlstrom JM Hagiwara K Finardi-Filho F Kjemtrup S Chrispeels MJ (1994) Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol 26: 1103-1113

Mohamed MF Coyne DP Read PE (1993) Shoot organogenesis in callus induced from pedicel explants of common bean (*Phaseolus vulgaris* L.). J Am Soc Hort Sci 118: 158-162

Mohamed MF Read PE Coyne DP (1992a) Plant regeneration from in vitro culture of embryogenic axis explants in common and tepary beans. J Am Soc Hort Sci 117: 332-336

Mohamed MF Read PE Coyne DP (1992b) Dark preconditioning, CPPU, and thidiazuron promote shoot organogenesis on seedling node explants of common and faba beans. J Am Soc Hort Sci 117: 668-672

Molvig L Tabe L Eggum B Moore A Craig S Spencer D Higgins TJV (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. Proc Natl Acad Sci USA 94: 8393-8398

Muhitch MJ (1997) Effects of expressing *E. coli* threonine synthase in tobacco (*Nicotiana tabacum* L.) suspension culture cells on free amino acid levels aspartate pathway enzyme activities and uptake of aspartate into the cells. J Plant Physiol 150: 16-22

Mühling M Gilroy J Croy RRD (1997) Legumin proteins from seeds of *Phaseolus vulgaris* L. J Plant Physiol 150: 489-492

Muntz K Veselin C Jung R Saalbach G Saalbach I Wadell D Pickardt T Schieder O (1997) Genetic engineering of high methionine proteins in grain legumes. In: Sulphur Metabolism in Higher Plants: Molecular, Ecophysiological and Nutritional Aspects (Cram WJ De Kok LJ Stulen I Brunold C Rennenberg H eds), Backhuys Publishers, Leiden, pp 71-86

Muntz K (1998) Deposition of storage proteins. Plant Mol Biol 38: 77-99

Muntz K Veselin C Saalbach G Saalbach I Waddell D Pickardt T Schieder O Wustenhagen T (1998) Genetic engineering for high methionine grain legumes. Nahrung 42: 125-127

Murashige T Skoog F (1962) Revised media for rapid growth and bio assays with tobacco tissue culture. Physiol Plant 15: 473-497

Murdock LL Huesing JE Nielsen SS Pratt RC Shade RE (1990) Biological effects of plant lectins on the cowpea weevil. Phytochemistry 29: 85-89

N

Nadolska-Orczyk A Orczyk W (2000) Study of the factors influencing *Agrobacterium*-mediated transformation of pea (*Pisum sativum* L.). Mol Breed 6: 185-194

Naito S Hirai Y Chino M Komeda Y (1994) Expression of a soybean (*Glycine max* [L] Merr) seed storage protein gene in transgenic *Arabidopsis thaliana* and its response to nutritional stress and to abscisic acid mutations. Plant Physiol 104: 497-503

Negrutiu I Cattoir-Reynaerts A Verbruggen I Jacobs M (1984) Lysine overproducer mutants with an altered dihydrodipicolinate synthase from protoplast culture of *Nicotiana sylvestris* (Spegazzini and Comes). Theor Appl Genet 68: 11-20

Negrutiu D De Brouwer D Dirks R Jacobs M (1985) Amino acid auxotrophs from protoplast cultures of *Nicotiana plumbaginifolia* Viviani. Mol Gen Genet 199: 330–337

Nelson OE (1969) Genetic modification of protein quality in plants. Adv Agron 21: 171-194

Neuhaus JM Rogers JC (1998) Sorting of proteins to vacuoles in plant cells Plant Mol Biol 38: 127-144

Nielsen NC Jung R Nam Y Beaman TW Oliveira LO Bassuner RB (1995) Synthesis and assembly of 11S globulins. J Plant Physiol 145: 641–647

Nishimura M Beevers H (1978) Hydrolases in vacuoles from castor bean endosperm. Plant Physiol 62: 44–48

Niu X Li X Veronese P Bressan RA Weller SC Hasegawa PM (2000) Factors affecting *Agrobacterium tumefaciens*-mediated transformation of peppermint. Plant Cell Rep 19: 304-310

Noji M Inoue K Kimura N Gouda A Saito K (1998) Isoform-dependent differences in feedback regulation and subcellular localization of serine acetyltransferase involved in cysteine biosynthesis from *Arabidopsis thaliana*. J Biol Chem 273: 32739-32745

Nordlee JA Taylor SL Townsend JA Thomas LA Bush RK (1996) Identification of a Brazil-nut allergen in transgenic soybean. New Engl J Med 334: 688-692

O

Odani S Odani S (1998) Isolation and primary structure of a methionine- and cystine-rich seed protein of *Cannabis sativa*. Biosci Biotechnol Biochem 62: 650-654

Okamuro JK Jofuku KD Goldberg RB (1986) Soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. Proc Natl Acad Sci USA 83: 8240-8244

Osborn TC Blake T Gepts P Bliss FA (1986) Bean arcelin 2: Genetic variation inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. Theor Appl Genet 71: 847-855

Osborn TC Burow M Bliss FA (1988) Purification and characterization of arcelin seed protein from common bean. Plant Physiol 86: 399-405

Osborne TB (1924) The vegetable proteins. Longmans Green & Company, London

P

Parker J (1995) World production of pulses trending upward Michigan. Dry Bean Digest 20: 2-5

Pedrazzini E Giovinazzo G Bielli A de Virgilio M Frigerio L Pesca M Faoro F Bollini R Ceriotti A Vitale A (1997) Protein quality control along the route to the plant vacuole. Plant Cell 9: 1869-1880

Pickardt T Saalbach I Waddell D Meixner M Müntz K Schieder O (1995) Seed specific expression of the 2S albumin gene from Brazil nut (*Bertholletia excelsa*) in transgenic *Vicia narbonensis*. Mol Breed 1: 295-301

Pickardt T Meixner M Waigand K Philipp J Demidov D Muntz K Galili G Schieder O (1998) Changes in the pools of free amino acids in the seeds of legumes: Expression of a deregulated *E. coli* aspartate kinase in seeds of *Vicia narbonensis*. Abstract presented at the International Eucarpia Symposium on breeding of protein and oil crops, april 1998, Pontevedra, Spain. Abstract book, pp 91-92

Piechocki MP Hines RN (1994) Oligonucleotide design and optimized protocol for site-directed mutagenesis. BioTechniques 16: 702-707

Pilon-Smits EHA Hwang S Lytle CM Zhu Y Tai JC Bravo RC Chen Y Leustek T Terry N (1999) Overexpression of ATP sulfurylase in Indian mustard leads to increased selenate uptake, reduction and tolerance. Plant Physiol 119: 123-132

Powers JR Whitaker JR (1977) Purification and some physical and chemical of red kidney bean (*Phaseolus vulgaris*) α-amylase inhibitor. J Food Biochem 1: 217-238

Pratt CR Nabhan PG (1988) Evolution and diversity of *Phaseolus acutifolius*. In: Genetic Resources of *Phaseolus* Bean (Gepts P ed), Kluwer Academic Publishers, Dordrecht, pp 409-440

Pusztai A Clarke EMW King TP (1979) The nutritional toxicity of *Phaseolus vulgaris* lectins. Proc Nutr Soc 38: 115-120

Pusztai A Grant G Stewart JC Bardocz S (1993) Nutritional evaluation of RAZ-2, a new *Phaseolus vulgaris* bean cultivar containing high levels of the natural insecticidal protein arcelin 1. J Agric Food Chem 41: 436-440

R

Raina A Datta A (1992) Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*. Proc Natl Acad Sci USA 89: 11774-11778

Randall PJ Thomson JA Schroeder ME (1979) Cotyledonary storage proteins in *Pisum sativum*. IV: Effect of sulfur phosphorus potassium and magnesium deficiencies. Aust J Plant Physiol 6: 11-24

Ravanel S Job D Douce R (1996) Purification and properties of cystathionine β-lyase from *Arabidopsis thaliana* overexpressed in *Escherichia coli*. Biochem J 320: 383–392

Ravanel S Gakière B Job D Douce R (1998) The specific features of methionine biosynthesis and metabolism in plants. Proc Natl Acad Sci USA 95: 7805–7812

Riesmeier J Klonus AK Pohlenz HD (1993) Purification to homogeneity and charaterization of homoserine kinase from wheat-germ. Phytochemistry 32: 581-583

Robinson DG Hoh B Hinz G Jeong BK (1995) One vacuole or two vacuoles: Do protein storage vacuoles arise de novo during pea cotyledon development? J Plant Physiol 145: 354-664

Romeo Andreas J Yandell BS Bliss FA (1986) Bean arcelin 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor Appl Genet 72: 123-128

Ruffet ML Droux M Douce R (1994) Purification and kinetic properties of serine acetyltransferase free of O-acetylserine(thiol)lyase from spinach chloroplast. Plant Physiol 140: 597-604

Russell DR Wallace KM Bathe JH Martinell BJ McCabe DE (1993) Stable transformation of *Phaseolus vulgaris* via electric-discharge mediated particle acceleration. Plant Cell Rep 12: 165-169

S

Saalbach G Christov V Jung R Saalbach I Manteuffel R Kunze G Bram-Barov K Müntz K (1995) Stable expression of vicilin from *Vicia faba* with eight additional single methionine residues but failure of accumulation of legumin with an attached peptide segment in tobacco seeds. Mol Breed 1: 245-258

Saito K Kurosawa M Tatsuguchi K Takagi Y Murakoshi I (1994) Modulation of cysteine biosynthesis in chloroplasts of transgenic tobacco overexpressing cysteine synthase [O-acetylserine (thiol)lyase]. Plant Physiol 106: 887–895

Saito K (1999) Biosynthesis of cysteine. In: Plant Amino Acids: Biochemistry and Biotechnology (Singh BK ed) Marcel Dekker, New York, pp 267–291

Saito K (2000) Regulation of sulfate transport and synthesis of sulfur-containing amino acids. Curr Opin Plant Biol 3: 188-195

Sanderfoot AA Raikhel NV (1999) The specificity of vesicle trafficking: Coat proteins and SNARs. Plant Cell 11: 629–641

Santino A Valsasina B Lioi L Vitale A Bollini R (1991) Bean (*Phaseolus vulgaris* L.) seed lectins: a novel electrophoretic variant of arcelin. Plant Physiol (Life Sci Adv) 10: 7-11

Schmidt A Jäger K (1992) Open questions about sulfur metabolism in plants. Annu Rev Plant Physiol Plant Mol Biol 43: 325–349

Schoonhoven AV Cardona C Valor J (1983) Resistance to the bean weevil and the Mexican bean weevil (Coleoptera: Bruchidae) in noncultivated common bean accessions. J Econ Entomol 76: 1255-1259

Schroeder HE (1982) Quantitative studies on the cotyledonary proteins in the genus *Pisum*. J Sci Food Agric 33: 623-633

Schroeder HE Gollasch S Moore A Tabe LM Craig S Hardie DC Chrispeels MJ Spencer D Higgins TJV (1995) Bean α-amylase inhibitor confers resistance to the pea weevil (*Bruchus pisorum*) in transgenic peas (*Pisum sativum* L.). Plant Physiol 107: 1233-1239

Scott ME Michaels TE (1992) *Xanthomonas* resistance of *Phaseolus* interspecific cross selections confirmed by field performance. HortScience 27: 348–350

Scrimshaw NS Waterlow JC Schürch B (1986) Energy and protein requirements. Eur J Clin Nutr 50: S1-52

Sengupta C DeLuca V Bailey DS Verma DPS (1981) Post-translational processing of 7S and 11S components of soybean storage proteins. Plant Mol Biol 1: 19-34

Sengupta-Gopalan C Reichert NA Barker RF Hall TC Kemp JD (1985) Developmentally regulated expression of the bean ß-phaseolin gene in tobacco seed. Proc Natl Acad Sci USA 82: 3320-3324

Sexton PJ Paek NC Shibles RM (1998) Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. Field Crops Res 59: 1-8

Sexton PJ Shibles RM (1999) Activity of ATP sulfurylase in reproductive soybean. Crop Sci 39: 131-135

Shaanan B Lis H Sharon N (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. Science 254: 862-866

Sharon N Lis H (1990) Legume lectins – a family of homologous proteins FASEB J 4: 3198-3208

Shaul O Galili G (1992) Increased lysine synthesis in transgenic tobacco plants expressing a bacterial dihydrodipicolinate synthase in their chloroplasts. Plant J 2: 203-209

Shaul O Galili G (1993) Concerted regulation of lysine and threonine synthesis in tobacco plants expressing bacterial feedback-insensitive aspartate kinase and dihydrodipicolinate synthase. Plant Mol Biol 23: 759-768

Shaw G Kamen R (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46: 659-667

Shen W-H Escudero J Schläppi M Ramos C Hohn B Koukolíková-Nicola Z (1993) T-DNA transfer to maize cells: Histochemical investigation of ß-glucuronidase activity in maize tissues. Proc Natl Acad Sci USA 90: 1488-1492

Shewry PR Faulks AJ Miflin BJ (1980) Effect of high-lysine mutations on the protein fractions of barley grain. Biochem Genet 18: 133-151

Shewry PR Tatham AS (1990) The prolamin storage proteins of cereal seeds: Structure and evolution. Biochem J: 267 1-12

Simon AE Tenbarge KM Scofield SR Finkelstein RR Crouch ML (1985) Nucleotide sequence of a cDNA clone of *Brassica napus* 12S storage protein shows homology with legumin from *Pisum sativum*. Plant Mol Biol 5: 191-201

Sindhu AS Zheng Z Murai N (1997) The pea seed storage protein legumin was synthesized processed and accumulated stably in transgenic rice endosperm. Plant Sci 130: 189-196

Singh R Axtell JD (1973) High lysine mutant gene (*hl*) that improves protein quality and biological value of grain sorghum. Crop Sci 13: 535-539

Singh SP (1999) Production and utilization. In: Common Bean Improvement in the Twenty-First Century (Singh SP ed) Kluwer Academic Publishers, Dordrecht, pp 1-24

Slightom JL Sun SM Hall TC (1983) Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc Natl Acad Sci USA 80: 1897-1901

Slightom JL Drong RF Klassy RC Hoffman LM (1985) Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucleic Acids Res 13: 6483-6498

Sotelo A (1996) Escumite bean (*Phaseolus acutifolius* A. Gray). In: Legumes and Oilseeds in Nutrition (Nwokolo E Smartt J eds) Chapman and Hall, London, pp 140-143

Spencer D Higgins TJV (1980) The biosynthesis of legumin in maturing pea seeds. Biochem Int 1: 502-509

Srivatanakul M Park SH Salas MG Smith RH (2001) Transformation of parameters enhancing T-DNA expression in kenaf (*Hibiscus cannabinus*). J Plant Physiol 158: 255-260

Stachel SE Messens E Van Montagu M Zambryski P (1985) Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. Nature 318: 624-629

Stachel SE Nester EW Zambryski PC (1986) A plant cell factor induces *Agrobacterium tumefaciens vir* gene expression. Proc Natl Acad Sci USA 83: 379-383

Staehelin LA (1997) The plant ER: A dynamic organelle composed of large number of discrete functional domains. Plant J 11: 1151–1165

Stöger E Parker M Christou P Casey R (2001) Pea Legumin overexpressed in Wheat endosperm assembles into an ordered paracrystalline matrix. Plant Physiol 125: 1732-1742

Sturm A Voelker TA Herman EM Chrispeels MJ (1988) Correct glycosylation Golgi-processing and targeting to protein bodies of the vacuolar protein phytohemagglutinin in transgenic tobacco. Planta 175: 170-183

Sun SSM (1999) Methionine enhancement in plants. In: Plant Amino Acids (Singh BK ed) Marcel Dekker, New York, pp 509-522

Sunilkumar G Rathore KS (2001) Transgenic cotton: factors influencing *Agrobacterium*-mediated transformation and regeneration. Mol Breed 8: 37-52

Suzuki K Ishimoto M Iwanaga M Kikuchi F Kitamura K (1995) Inheritance of seed α-amylase inhibitor in the common bean and genetic relationship to arcelin. Theor Appl Genet 90: 762-766

Suzuki S Nakano M (2002) *Agrobacterium*-mediated production of transgenic plants of *Muscari armeniacum* Leichtl ex Bak. Plant Cell Rep 20: 835-841

T

Tabe LM Wardley-Richardson T Ceriotti A Aryan A McNabb W Moore A Higgins TJV (1995) Technological approach to improving the nutritive value of alfalfa. J Anim Sci 73: 2752-2759

Tabe L Higgins TJV (1998) Engineering plant protein composition for improved nutrition. Trends Plant Sci 3: 282-286

Tabe LM Droux M (2001) Sulfur assimilation in developing lupin cotyledons could contribute significantly to the accumulation of organic sulfur reserves in the seed. Plant Physiol 126: 176-187

Tabe L Droux M (2002) Limits to sulfur accumulation in transgenic lupin seeds expressing a foreign sulfur-rich protein. Plant Physiol 128: 1137-1148

Takahashi H Yamazaki M Sasakura N Watanabe A Leustek T de Almeida Engler J Engler G Van Montagu M Saito K (1997) Regulation of sulfur assimilation in higher plants: A sulfate transporter induced in sulfate-starved roots plays a central role in *Arabidopsis thaliana*. Proc Natl Acad Sci 94: 11102–11107

Takaiwa F Katsube T Kitagawa S Hisago Y Kito M Utsumi S (1995) High level accumulation of soybean in vacuole-derived protein bodies in the endosperm tissue of transgenic tobacco seed. Plant Sci 111: 39-49

Takasaki T Hatakeyama K Ojima K Watanabe M Toriyama K Hinata K (1997) Factors influencing *Agrobacterium*-mediated transformation of *Brassica rapa* L. Breeding Sci 47: 127-134

Tang G Zhu-Shimoni JX Amir R Zchori IB Galili G (1997) Cloning and expression of an *Arabidopsis thaliana* cDNA encoding a monofunctional aspartate kinase homologous to the lysine-sensitive enzyme of *Escherichia coli*. Plant Mol Biol 34: 287-293

Thoen A Rognes SE Aarnes H (1978) Biosynthesis of threonine from homoserine in pea seedlings. 1: Homoserine kinase. Plant Sci Lett 13: 103-112

Thomas TL (1993) Gene expression during plant embryogenesis and germination: An overview. Plant Cell 5: 1401-1410

Thompson GA Datko AH Mudd SH Giovanelli J (1982) Methionine biosynthesis in *Lemna:* Inhibition of cystathionine γ-synthase O-phospho-homoserine sulfhydrolase and O-acetylserine sulfhydrolase. Plant Physiol 69: 1077–1083

Thompson JF Madison JT (1990) The effect of sulphate and methionine on legume proteins. In: Sulfur Nutrition and Sulfur Assimilation in Higher Plants (Rennenberg H Brunold C De Kok LJ Stulen I eds) SPB Academic Publishers, Den Haag

Townsend JF Madison JT (1994) Factors which influence the *Agrobacterium*-mediated transformation of soybean. J Cell Biochem Suppl 18A: X1-014

Trifonova A Madsen S Olesen A (2001) *Agrobacterium*-mediated transgene delivery and integration into barley under a range of *in vitro* culture conditions. Plant Sci 161: 871-880

Trossat C Nolte KD Hanson AD (1996) Evidence that the pathway of dimethylsulfoniopropionate biosynthesis begins in the cytosol and ends in the chloroplast. Plant Physiol 111: 965-973

Tu HM Godfrey LW Sun SSM (1998) Expression of the Brazil nut methionine-rich protein and mutants with increased methionine in transgenic potato. Plant Mol Biol 37: 829-838

Turk SCHJ Melchers LS den Dulk-Ras H Regensburg-Tuïnk AJG Hooykaas PJJ (1991) Environmental conditions differentially affect *vir* gene induction in different *Agrobacterium* strains. Role of the VirA sensor protein. Plant Mol Biol 16: 1051-1059

Turner WL Pallett KE Lea PJ (1998) Cystathionine β-lyase from *Echinochloa colonum* tissue culture. Phytochemistry 47: 189–196

U

Ueng P Galili G Sapanara V Goldsbrough PB Dube P Beachy RN Larkins BA (1988) Expression of a maize storage protein gene in petunia plants is not restricted to seeds. Plant Physiol 86: 1281-1285

Utsumi S Kitagawa S Katsube T Kang IJ Gidamis AB Takaiwa F Kito M (1993) Synthesis, processing and accumulation of modified glycinins of soybean in the seeds leaves and stems of transgenic tobacco. Plant Sci 92: 191-202

V

Valvekens D Van Montagu M Van Lijsebettens M (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis* root explants using kanamycin selection. Proc Natl Acad Sci USA 85: 5536-5540

van Aarssen R Soetaert P Stam M Dockx J Gosselé V Seurinck J Reynaerts A Cornelissen M (1995) *cry* IA(b) transcript formation in tobacco is inefficient. Plant Mol Biol 28: 513-524

Van der Wilden W Herman EM Chrispeels MJ (1980) Protein bodies of mung bean cotyledons as autophagic organelles. Proc Natl Acad Sci USA 77: 428–432

Vancanneyt G Schmidt R O'Connor-Sanchez A Willmitzer L Rocha-Sosa M (1990) Construction of an intron-containing marker gene: Splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. Mol Gen Genet 220: 245-250

Vaquero F Robles C Ruiz ML (1993) A method for long-term micropropagation of *Phaseolus coccineus* L. Plant Cell Rep 12: 395–398

Vitale A Denecke J (1999) The endoplasmic reticulum—Gateway of the secretory pathway. Plant Cell 11: 615–628

von Heijne G (1984) How signal sequences maintain cleavage specificity. J Mol Biol 173: 243–251

W

Waddell DR Saalbach I Pickardt T enneking D Fleckenstein J Schnug E Muntz K (1997) Effect of over-expression of a sulphur-rich 2S albumin on the sulphur metabolism of seed in transgenic *Vicia narbonensis*. Abstract presented at the 7th International Seed Protein Symposium, June 24-26, Gaterleben, Germany.

Wallsgrove RM Lea PJ Miflin BJ (1983) Intracellular localization of aspartate kinase and the enzymes of threonine and methionine biosynthesis in green leaves. Plant Physiol 71: 780-784

Wang ZY Ye XD Nagel J Potrykus I Spangenberg G (2001) Expression of a sulphur-rich sunflower albumin gene in transgenic tall fescue (*Festuca arundinacea* Schreb.) plants. Plant Cell Rep 20: 213-219

White CL Tabe LM Dove H Hamblin J Young NP Taylor R Gulati S Ashes J Higgins TJV (2000) Increased efficiency of wool growth and live weight gain in Merino sheep fed transgenic lupin seed containing sunflower albumin. J Sci Food Agric 81: 147-154

Wilson BJ Gray AC Matthews BF (1991) Bifunctional protein in carrot contains both aspartokinase and homoserine dehydrogenase activities. Plant Physiol 97: 1323-1328

Wright DJ (1988) In: Developments in Food Proteins, Vol 6 (Development Series) (Hudson BJF ed) Applied Science, London, pp 119-177

Wrigley CW DuCross DL Archer MJ Downie PG Roxburgh CM (1980) The sulphur content of wheat endosperm proteins and its relevance to grain quality. Aust J Plant Physiol 7: 755-766

Y

Yet MB Shao MC Wold F (1988) Effects of the protein matrix on glycan processing in glycoproteins. FASEB J 2: 22-31

Young NM Thibault P Watson DC Chrispeels M (1999) Post-translational processing of two α-amylase inhibitors and an arcelin from common bean *Phaseolus vulgaris*. FEBS Lett 446: 203-206

Young NM Watson DC Yaguchi M Adar R Arango R Rodriguez-Arango E Sharon N Blay PKS Thibault P (1995) C-terminal post-translational proteolysis of plant lectins and their recombinant forms expressed in *Escherichia coli*. Characterization of "ragged ends" by mass spectrometry. J Biol Chem 270: 2563-2570

Z

Zambre M De Clercq J Vranová E Van Montagu M Angenon G Dillen W (1998a) Plant regeneration from embryo-derived callus in *Phaseolus vulgaris* L. (common bean) and *P. acutifolius* A. Gray (tepary bean). Plant Cell Rep 17: 626-630

Zambre M De Clercq J Van Montagu M Angenon G Dillen W (1998b) Shoot regeneration from callus in wild genotypes of *Phaseolus vulgaris* L. Ann Rep Bean Improv Coop 41: 119-120

Zambre M Geerts P Maquet A Van Montagu M Dillen W Angenon G (2001) Regeneration of fertile plants from callus in *Phaseolus polyanthus* Greenman (Year Bean). Ann Bot 88: 371-377

Zambre M Terryn N De Clercq J De Buck S Dillen W Van Montagu M Van Der Straeten D Angenon G (2002) Light strongly promotes gene transfer from *Agrobacterium tumefaciens* to plant cells. Planta: in press

Zeh M Casazza AP Kreft O Roessner U Bieberich K Willmitzer L Hoefgen R Hesse H (2001) Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. Plant Physiol 127: 792-802

Zhang FL Yakahata Y Watanabe M Xu JB (2000) *Agrobacterium*-mediated transformation of cotyledonary explants of Chinese cabbage (*Brassica campestris* L. ssp *pekinensis*). Plant Cell Rep 19: 569-575

Zhang Z Coyne DP Mitra A (1997) Factors affecting *Agrobacterium*-mediated transformation of common bean. J Am Soc Hort Sci 122: 300-305

Zheng Z Sumi K Tanaka K Murai N (1995) The bean seed storage protein ß-phaseolin is synthesized processed and accumulated in the vacuolar type-II protein bodies of transgenic rice endosperm. Plant Physiol 109: 777-786

Zhu-Shimoni JX Lev-Yadun S Matthews B Galili G (1997) Expression of an aspartate kinase homoserine dehydrogenase gene is subject to specific spatial and temporal regulation in vegetative tissues, flowers, and developing seeds. Plant Physiol 113: 695-706

Zuo WN Sun SSM (1996) Purification and characterization of the methionine-rich 2S albumin seed proteins from the Brazil nut family. J Agric Food Chem 44: 1206-1210